

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
19 September 2002 (19.09.2002)

PCT

(10) International Publication Number
WO 02/072596 A1

(51) International Patent Classification⁷: C07H 21/02, 21/04, C12P 21/06, C12N 5/00

(21) International Application Number: PCT/US02/07053

(22) International Filing Date: 7 March 2002 (07.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/802,520 9 March 2001 (09.03.2001) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): LAL, Preeti, G. [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). FARIS, Mary [US/US]; 2538 Almaden Court, Los Angeles, CA 90077 (US). CHEN, Huei-Mei [US/US]; 2425 Lyle Court, Santa Clara, CA 95056 (US). ISON, Craig, H. [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: STREETER, David, G. et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/072596 A1

(54) Title: STEAP-RELATED PROTEIN

(57) Abstract: The invention provides a cDNA which encodes a STEAP-related protein. It also provides for the use of the cDNA, fragments, complements, and variants thereof and of the encoded protein, portions thereof and antibodies thereto for diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention additionally provides expression vectors and host cells for the production of the protein and a transgenic model system.

STEAP-RELATED PROTEIN**TECHNICAL FIELD**

5 This invention relates to cDNA which encodes a STEAP-related protein and to the use of the cDNA and the encoded protein in the diagnosis and treatment of prostate cell proliferative disorders, in particular, prostate hyperplasia and prostate cancer.

BACKGROUND OF THE INVENTION

Phylogenetic relationships among organisms have been demonstrated many times, and studies 10 from a diversity of prokaryotic and eukaryotic organisms suggest a more or less gradual evolution of molecules, biochemical and physiological mechanisms, and metabolic pathways. Despite different evolutionary pressures, the proteins of nematode, fly, rat, and man have common chemical and structural features and generally perform the same cellular function. Comparisons of the nucleic acid and protein sequences from organisms where structure and/or function are known accelerate the 15 investigation of human sequences and allow the development of model systems for testing diagnostic and therapeutic agents for human conditions, diseases, and disorders.

Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate 20 cancer and more than 33,000 deaths from the disorder each year. Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

25 As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain, or lung. About 80% of these tumors remain responsive to 30 androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is currently no known treatment for this condition.

A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer

exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists and clinicians maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

5 Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role in tumor development, growth, and progression. The growth factors epidermal growth factor (EGF), fibroblast growth factor (FGF), and transforming growth factor alpha (TGF α) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of
10 tumor development and progression, and affect signaling pathways in these cells in various ways (Lin *et al.* (1999) *Cancer Res* 59:2891-2897; Putz *et al.* (1999) *Cancer Res* 59:227-233). The TGF- β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival
(Gold (1999) *Crit Rev Oncog* 10:303-360). Finally, there are human cell lines representing both the
15 androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung (1999) *Prostate* 38:199-207).

20 Six-transmembrane epithelial antigen of the prostate (STEAP) is a prostate-specific cell-surface marker (Hubert *et al.* (1999) *Proc Natl Acad Sci* 96:14523-14528). STEAP is 339 amino acids in length and has six predicted membrane-spanning regions. It is highly expressed in normal and cancerous prostate tissues and in several prostate cancer-derived cell lines. Its level of expression is insensitive to the presence of androgen. Immunostaining shows that STEAP is located at the plasma membrane of prostate cells where it concentrates at cell-cell junctions of the secretory epithelium.
25 Cell surface antigens such as STEAP may be useful in antibody therapy, cancer-vaccines, and diagnostic imaging for treatment of prostate cancer.

The discovery of a cDNA encoding STEAP-related protein satisfies a need in the art by providing compositions which are useful in the diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

30

SUMMARY OF THE INVENTION

The invention is based on the discovery of a cDNA encoding STEAP-related protein (STEAPRP) which is useful in the diagnosis and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

35 The invention provides an isolated cDNA comprising a nucleic acid sequence encoding a protein having the amino acid sequence of SEQ ID NO:1. The invention also provides an isolated

cDNA or the complement thereof selected from the group consisting of a nucleic acid sequence of SEQ ID NO:2, a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-9, and a variant of SEQ ID NO:2, SEQ ID NO:10. The invention additionally provides a composition, a substrate, and a probe comprising the cDNA, or the complement of the cDNA, encoding STEAPRP. The invention further 5 provides a vector containing the cDNA, a host cell containing the vector and a method for using the cDNA to make STEAPRP. The invention still further provides a transgenic cell line or organism comprising the vector containing the cDNA encoding STEAPRP. The invention additionally provides a fragment, a variant, or the complement of the cDNA selected from the group consisting of SEQ ID Nos:2-10. In one aspect, the invention provides a substrate containing at least one of these fragments 10 or variants or the complements thereof. In a second aspect, the invention provides a probe comprising a cDNA or the complement thereof which can be used in methods of detection, screening, and purification. In a further aspect, the probe is a single-stranded complementary RNA or DNA molecule.

The invention provides a method for using a cDNA to detect the differential expression of a 15 nucleic acid in a sample comprising hybridizing a probe to the nucleic acids, thereby forming hybridization complexes and comparing hybridization complex formation with a standard, wherein the comparison indicates the differential expression of the cDNA in the sample. In one aspect, the method of detection further comprises amplifying the nucleic acids of the sample prior to hybridization. In another aspect, the method showing differential expression of the cDNA is used to 20 diagnose prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. In another aspect, the cDNA or a fragment or a variant or the complements thereof may comprise an element on an array.

The invention additionally provides a method for using a cDNA or a fragment or a variant or the complements thereof to screen a library or plurality of molecules or compounds to identify at least 25 one ligand which specifically binds the cDNA, the method comprising combining the cDNA with the molecules or compounds under conditions allowing specific binding, and detecting specific binding to the cDNA, thereby identifying a ligand which specifically binds the cDNA. In one aspect, the molecules or compounds are selected from aptamers, DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and 30 regulatory molecules.

The invention provides a purified protein or a portion thereof selected from the group consisting of an amino acid sequence of SEQ ID NO:1, a variant having at least 55% identity to the amino acid sequence of SEQ ID NO:1, an antigenic epitope of SEQ ID NO:1, and a biologically active portion of SEQ ID NO:1. The invention also provides a composition comprising the purified protein 35 in conjunction with a pharmaceutical carrier. The invention further provides a method of using the

STEAPRP to treat a subject with prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer comprising administering to a patient in need of such treatment the composition containing the purified protein. The invention still further provides a method for using a protein to screen a library or a plurality of molecules or compounds to identify at least one ligand, the method 5 comprising combining the protein with the molecules or compounds under conditions to allow specific binding and detecting specific binding, thereby identifying a ligand which specifically binds the protein. In one aspect, the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs. In another aspect, the ligand is used to treat a subject with 10 prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

The invention provides a method of using a protein to screen a subject sample for antibodies which specifically bind the protein comprising isolating antibodies from the subject sample, contacting the isolated antibodies with the protein under conditions that allow specific binding, dissociating the antibody from the bound-protein, and comparing the quantity of antibody with known 15 standards, wherein the presence or quantity of antibody is diagnostic of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

The invention also provides a method of using a protein to prepare and purify antibodies comprising immunizing a animal with the protein under conditions to elicit an antibody response, isolating animal antibodies, attaching the protein to a substrate, contacting the substrate with isolated 20 antibodies under conditions to allow specific binding to the protein, dissociating the antibodies from the protein, thereby obtaining purified antibodies.

The invention provides a purified antibody which binds specifically to a protein which is expressed in prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention also provides a method of using an antibody to diagnose prostate cell proliferative 25 disorders, particularly prostate hyperplasia and prostate cancer comprising combining the antibody comparing the quantity of bound antibody to known standards, thereby establishing the presence of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. The invention further provides a method of using an antibody to treat prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer comprising administering to a patient in need of 30 such treatment a pharmaceutical composition comprising the purified antibody.

The invention provides a method for inserting a heterologous marker gene into the genomic DNA of a mammal to disrupt the expression of the endogenous polynucleotide. The invention also provides a method for using a cDNA to produce a mammalian model system, the method comprising constructing a vector containing the cDNA selected from SEQ ID NOs:2-10, transforming the vector 35 into an embryonic stem cell, selecting a transformed embryonic stem, microinjecting the transformed

embryonic stem cell into a mammalian blastocyst, thereby forming a chimeric blastocyst, transferring the chimeric blastocyst into a pseudopregnant dam, wherein the dam gives birth to a chimeric offspring containing the cDNA in its germ line, and breeding the chimeric mammal to produce a homozygous, mammalian model system.

5

BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G show the STEAPRP (SEQ ID NO:1) encoded by the cDNA (SEQ ID NO:2). The translation was produced using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA).

Figures 2A, 2B, and 2C demonstrate the conserved chemical and structural similarities among 10 the sequences and domains of STEAPRP (7492448; SEQ ID NO:1) and human STEAP (g6572948; SEQ ID NO:11). The alignment was produced using the MEGALIGN program of LASERGENE software (DNASTAR, Madison WI).

Tables 1 and 2 show the northern analysis for STEAPRP produced using the LIFESEQ Gold database (Incyte Genomics, Palo Alto CA). In Table 1, the first column presents the tissue categories; 15 the second column, the total number of clones in the tissue category; the third column, the ratio of the number of libraries in which at least one transcript was found to the total number of libraries; the fourth column, absolute clone abundance of the transcript; and the fifth column, percent abundance of the transcript. Table 2 shows expression of STEAPRP in prostate tissues, particularly from patients with cancer. The first column lists the library name, the second column, the number of clones 20 sequenced for that library; the third column, the description of the tissue from which the library was derived; the fourth column, the absolute abundance of the transcript; and the fifth column, the percent abundance of the transcript.

Table 3 shows the differential expression of STEAPRP in human LNCaP prostate carcinoma cells compared to human PrEC nontumorigenic prostate epithelial cells as determined by microarray 25 analysis. Column 1 lists the mean differential expression (DE) values presented as log2 DE (LNCaP cells/PrEC cells). Column 2 lists the percentage covariance (CV%) in differential expression values. Column 3 lists the PrEC-derived samples labeled with fluorescent green dye Cy3. Column 4 lists the LNCaP-derived samples labeled with fluorescent red dye Cy5.

DESCRIPTION OF THE INVENTION

30 It is understood that this invention is not limited to the particular machines, materials and methods described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments and is not intended to limit the scope of the present invention which will be limited only by the appended claims. As used herein, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. For example, a reference 35 to "a host cell" includes a plurality of such host cells known to those skilled in the art.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"STEAPRP" refers to a purified protein obtained from any mammalian species, including bovine, canine, murine, ovine, porcine, rodent, simian, and preferably the human species, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

"Array" refers to an ordered arrangement of at least two cDNAs on a substrate. At least one of the cDNAs represents a control or standard, and the other, a cDNA of diagnostic or therapeutic interest. The arrangement of from about two to about 40,000 cDNAs on the substrate assures that the size and signal intensity of each labeled hybridization complex formed between each cDNA and at least one sample nucleic acid is individually distinguishable.

The "complement" of a cDNA of the Sequence Listing refers to a nucleic acid molecule which is completely complementary over its full length and which will hybridize to the cDNA or an mRNA under conditions of maximal stringency.

"cDNA" refers to an isolated polynucleotide, nucleic acid molecule, or any fragment or complement thereof. It may have originated recombinantly or synthetically, may be double-stranded or single-stranded, represents coding and noncoding 3' or 5' sequence, and generally lacks introns.

The phrase "cDNA encoding a protein" refers to a nucleotide sequence that closely aligns with sequences which encode conserved regions, motifs or domains that were identified by employing analyses well known in the art. These analyses include BLAST (Basic Local Alignment Search Tool) which provides identity within the conserved region (Altschul (1993) J Mol Evol 36: 290-300; Altschul et al. (1990) J Mol Biol 215:403-410).

A "composition" comprises the polynucleotide and a labeling moiety or a purified protein in conjunction with a pharmaceutical carrier.

"Derivative" refers to a cDNA or a protein that has been subjected to a chemical modification. Derivatization of a cDNA can involve substitution of a nontraditional base such as queosine or of an analog such as hypoxanthine. These substitutions are well known in the art. Derivatization of a protein involves the replacement of a hydrogen by an acetyl, acyl, alkyl, amino, formyl, or morpholino group. Derivative molecules retain the biological activities of the naturally occurring molecules but may confer advantages such as longer lifespan or enhanced activity.

"Differential expression" refers to an increased, upregulated or present, or decreased,

downregulated or absent, gene expression as detected by presence, absence or at least two-fold changes in the amount of transcribed messenger RNA or translated protein in a sample.

"Disorder" refers to conditions, diseases or syndromes in which the cDNAs and STEAPRP are differentially expressed. Such a disorder includes prostate cell proliferative disorders, particularly 5 prostate hyperplasia and prostate cancer.

"Fragment" refers to a chain of consecutive nucleotides from about 50 to about 4000 base pairs in length. Fragments may be used in PCR or hybridization technologies to identify related nucleic acid molecules and in binding assays to screen for a ligand. Such ligands are useful as therapeutics to regulate replication, transcription or translation.

10 A "hybridization complex" is formed between a cDNA and a nucleic acid of a sample when the purines of one molecule hydrogen bond with the pyrimidines of the complementary molecule, e.g., 5'-A-G-T-C-3' base pairs with 3'-T-C-A-G-5'. Hybridization conditions, degree of complementarity and the use of nucleotide analogs affect the efficiency and stringency of hybridization reactions.

15 "Labeling moiety" refers to any visible or radioactive label than can be attached to or incorporated into a cDNA or protein. Visible labels include but are not limited to anthocyanins, green fluorescent protein (GFP), β glucuronidase, luciferase, Cy3 and Cy5, and the like. Radioactive markers include radioactive forms of hydrogen, iodine, phosphorous, sulfur, and the like.

20 "Ligand" refers to any agent, molecule, or compound which will bind specifically to a polynucleotide or to an epitope of a protein. Such ligands stabilize or modulate the activity of polynucleotides or proteins and may be composed of inorganic and/or organic substances including minerals, cofactors, nucleic acids, proteins, carbohydrates, fats, and lipids.

25 "Oligonucleotide" refers a single-stranded molecule from about 18 to about 60 nucleotides in length which may be used in hybridization or amplification technologies or in regulation of replication, transcription or translation. Substantially equivalent terms are amplimer, primer, and oligomer.

"Portion" refers to any part of a protein used for any purpose; but especially, to an epitope for the screening of ligands or for the production of antibodies.

30 "Post-translational modification" of a protein can involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and the like. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cellular location, cell type, pH, enzymatic milieu, and the like.

35 "Probe" refers to a cDNA that hybridizes to at least one nucleic acid in a sample. Where targets are single-stranded, probes are complementary single strands. Probes can be labeled with reporter molecules for use in hybridization reactions including Southern, northern, in situ, dot blot, array, and like technologies or in screening assays.

"Protein" refers to a polypeptide or any portion thereof. A "portion" of a protein refers to that length of amino acid sequence which would retain at least one biological activity, a domain identified by PFAM or PRINTS analysis or an antigenic epitope of the protein identified using Kyte-Doolittle algorithms of the PROTEAN program (DNASTAR, Madison WI). An "oligopeptide" is an amino acid sequence from about five residues to about 15 residues that is used as part of a fusion protein to produce an antibody.

"Purified" refers to any molecule or compound that is separated from its natural environment and is from about 60% free to about 90% free from other components with which it is naturally associated.

10 "Sample" is used in its broadest sense as containing nucleic acids, proteins, antibodies, and the like. A sample may comprise a bodily fluid; the soluble fraction of a cell preparation, or an aliquot of media in which cells were grown; a chromosome, an organelle, or membrane isolated or extracted from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; a cell; a tissue; a tissue print; a fingerprint, buccal cells, skin, or hair; and the like.

15 "Specific binding" refers to a special and precise interaction between two molecules which is dependent upon their structure, particularly their molecular side groups. For example, the intercalation of a regulatory protein into the major groove of a DNA molecule or the binding between an epitope of a protein and an agonist, antagonist, or antibody.

"Similarity" as applied to sequences, refers to the quantification (usually percentage) of 20 nucleotide or residue matches between at least two sequences aligned using a standardized algorithm such as Smith-Waterman alignment (Smith and Waterman (1981) J Mol Biol 147:195-197) or BLAST2 (Altschul *et al.* (1997) Nucleic Acids Res 25:3389-3402). BLAST2 may be used in a standardized and reproducible way to insert gaps in one of the sequences in order to optimize alignment and to achieve a more meaningful comparison between them. Particularly in proteins, 25 similarity is greater than identity in that conservative substitutions, for example, valine for leucine or isoleucine, are counted in calculating the reported percentage. Substitutions which are considered to be conservative are well known in the art.

"Substrate" refers to any rigid or semi-rigid support to which cDNAs or proteins are bound and includes membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, 30 capillaries or other tubing, plates, polymers, and microparticles with a variety of surface forms including wells, trenches, pins, channels and pores.

"Variant" refers to molecules that are recognized variations of a cDNA or a protein encoded by the cDNA. Splice variants may be determined by BLAST score, wherein the score is at least 100, and most preferably at least 400. Allelic variants have a high percent identity to the cDNAs and may 35 differ by about three bases per hundred bases. "Single nucleotide polymorphism" (SNP) refers to a

change in a single base as a result of a substitution, insertion or deletion. The change may be conservative (purine for purine) or non-conservative (purine to pyrimidine) and may or may not result in a change in an encoded amino acid or its secondary, tertiary, or quaternary structure.

THE INVENTION

5 The invention is based on the discovery of a cDNA which encodes STEAPRP and on the use of the cDNA, or fragments thereof, and protein, or portions thereof, directly or as compositions in the characterization, diagnosis, and treatment of prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

Nucleic acids encoding the STEAPRP of the present invention were first identified in Incyte 10 Clone 7100809 from the brain dentate nucleus cDNA library (BRAWTDR02) using a computer search for nucleotide and/or amino acid sequence alignments. SEQ ID NO:2 (7492448CB1) was derived from the following overlapping and/or extended nucleic acid sequences (SEQ ID NO:3-9): Incyte Clones 7100809H1 (BRAWTDR02), 6912820J1 (PITUDIR01), 4647117F6 (PROSTUT20), 15 7004364H1 (COLNFEC01), 70351677D1 (SG0000177), 4108079H1 (PROSBPT07), and 4669848H1 (SINTNOT24). Tables 1 shows expression of the transcript across the tissue categories, and the highest abundance of the transcript is found in male reproductive tissues (42%). STEAPRP is expressed exclusively in prostate tissue in this category. Table 2 shows expression of the transcript in prostate tissues, particularly in tissues from patients with adenofibromatous hyperplasia, prostate intraepithelial neoplasia, and adenocarcinoma. STEAPRP is expressed in prostate tissue libraries 20 (PROSNOT19, PROSDIT01, PROSNOT20, and PROSNOT06) from patients with adenofibromatous hyperplasia, a prostate tissue library (PROETMP06) from a patient with intraepithelial neoplasia, and prostate tissue libraries (PROSTUT18, PROSTUS20, PROSTUT04, PROSTUT21, PROSTUS19, and PROSTUT12) from patients with adenocarcinoma. Table 3 shows the differential expression of STEAPRP in human LNCaP prostate carcinoma cells compared to PrEC nontumorigenic prostate 25 epithelial cells. Cells were grown under different conditions in the experiments. Starved cells were grown in basal media in the absence of growth factors and hormones. Rich media contained growth factors and nutrients to promote growth. STEAPRP shows increased expression in LNCaP carcinoma cells relative to PrEC under all growth conditions. The transcript is therefore useful in diagnostic assays for prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer. A 30 fragment of the cDNA from about nucleotide 1 to about nucleotide 50 is also useful in diagnostic assays.

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1 as shown in Figures 1A, 1B, 1C, 1D, 1E, 1F, and 1G. STEAPRP is 490 amino acids in length and has one potential N-glycosylation site at N256; one potential cyclic AMP- 35 or cyclic GMP-dependent protein kinase phosphorylation site at T32; six potential casein kinase II

phosphorylation sites at S12, T77, S100, S128, S197, and S348; five potential protein kinase C phosphorylation sites at S9, T46, S197, S328, and S455; and one potential tyrosine kinase phosphorylation site at Y423. PFAM analysis indicates that the region of STEAPRP from T32 to L136 is similar to a KTN NAD-binding domain. KTN NAD-binding domains are found in a variety 5 of proteins, including potassium channels, phosphoesterases, and various transporters. BLOCKS analysis indicates that the region of STEAPRP from G34 to K64 is similar to bacterial-type phytoene dehydrogenase, the region from T32 to V56 is similar to pyridine nucleotide-disulfide class II oxidoreductases, and the region from T32 to F70 is similar to 6-phosphogluconate dehydrogenase. PRINTS analysis indicates that the region of STEAPRP from V317 to Y331 is similar to a phthalate 10 dioxygenase reductase family signature and the region from V33 to I47 is similar to an adrenodoxin reductase family signature. The presence of these motifs indicates a possible function for STEAPRP in oxido-reductase reactions. Hidden Markov Model analysis of STEAPRP indicates the presence of six transmembrane regions from T210 to P238, from E253 to Q281, from C301 to S328, from M359 to I379, from F391 to L411, and from F426 to I454; and the presence of a signal peptide region from 15 M359 to N387. As shown in Figures 2A, 2B and 2C, STEAPRP has chemical and structural similarity with human STEAP (g6572948; SEQ ID NO:11). In particular, STEAPRP and STEAP share about 43% identity and the six predicted transmembrane regions. Useful antigenic epitopes extend from about G59 to about D75, from about D234 to about K249, and from about S455 to about T478; and a biologically active portion of STEAPRP extends from about T32 to about L136. An antibody which 20 specifically binds STEAPRP is useful in a diagnostic assay to identify prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

Mammalian variants of the cDNA encoding STEAPRP were identified using BLAST2 with default parameters and the ZOOSEQ databases (Incyte Genomics). These preferred variants have about 85% identity as shown in the table below. The first column shows the SEQ ID for the human 25 cDNA (SEQ ID_H); the second column, the SEQ ID for the variant cDNAs (SEQ ID_{var}); the third column, the clone number for the variant cDNAs (Clone_{var}); the fourth column, the library name; the fifth column, the alignment of the variant cDNA to the human cDNA; and the sixth column, the percent identity to the human cDNA.

30	SEQ	SEQ ID _H	SEQ ID _{var}	Clone _{var}	Library Name	Nt _H Alignment	Identity
		2	10	702819778T1	RATSNON03	285-607	85%

The cDNA, SEQ ID NO:10 is particularly useful for producing transgenic cell lines or organisms. 35 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of cDNAs encoding STEAPRP, some bearing minimal similarity to the cDNAs of

any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of cDNA that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide encoding naturally occurring STEAPRP, and all such variations are to 5 be considered as being specifically disclosed.

The cDNAs of SEQ ID NOS:2-10 may be used in hybridization, amplification, and screening technologies to identify and distinguish among SEQ ID NO:2 and related molecules in a sample. The mammalian cDNAs may be used to produce transgenic cell lines or organisms which are model systems for human prostate cell proliferative disorders, particularly prostate hyperplasia and prostate 10 cancer and upon which the toxicity and efficacy of potential therapeutic treatments may be tested.

Toxicology studies, clinical trials, and subject/patient treatment profiles may be performed and monitored using the cDNAs, proteins, antibodies and molecules and compounds identified using the cDNAs and proteins of the present invention.

Characterization and Use of the Invention

15 cDNA libraries

In a particular embodiment disclosed herein, mRNA is isolated from mammalian cells and tissues using methods which are well known to those skilled in the art and used to prepare the cDNA libraries. The Incyte cDNAs were isolated from mammalian cDNA libraries prepared as described in the EXAMPLES. The consensus sequences are chemically and/or electronically assembled from 20 fragments including Incyte cDNAs and extension and/or shotgun sequences using computer programs such as PHRAP (P Green, University of Washington, Seattle WA), and AUTOASSEMBLER application (Applied Biosystems, Foster City CA). After verification of the 5' and 3' sequence, at least one representative cDNA which encodes STEAPRP is designated a reagent.

Sequencing

25 Methods for sequencing nucleic acids are well known in the art and may be used to practice any of the embodiments of the invention. These methods employ enzymes such as the Klenow fragment of DNA polymerase I, SEQUENASE, Taq DNA polymerase and thermostable T7 DNA polymerase (Amersham Pharmacia Biotech (APB), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 system (Hamilton, Reno NV) and the DNA ENGINE thermal cycler (MJ Research, Watertown MA). Machines commonly used for sequencing include the ABI PRISM 3700, 377 or 373 DNA sequencing systems (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (APB), and the like. The sequences may be analyzed using a variety of algorithms well known 30 in the art and described in Ausubel *et al.* (1997; Short Protocols in Molecular Biology, John Wiley & 35

Sons, New York NY, unit 7.7) and in Meyers (1995; Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

Shotgun sequencing may also be used to complete the sequence of a particular cloned insert of interest. Shotgun strategy involves randomly breaking the original insert into segments of various sizes and cloning these fragments into vectors. The fragments are sequenced and reassembled using overlapping ends until the entire sequence of the original insert is known. Shotgun sequencing methods are well known in the art and use thermostable DNA polymerases, heat-labile DNA polymerases, and primers chosen from representative regions flanking the cDNAs of interest. Incomplete assembled sequences are inspected for identity using various algorithms or programs such as CONSED (Gordon (1998) *Genome Res* 8:195-202) which are well known in the art.

Contaminating sequences, including vector or chimeric sequences, or deleted sequences can be removed or restored, respectively, organizing the incomplete assembled sequences into finished sequences.

Extension of a Nucleic Acid Sequence

The sequences of the invention may be extended using various PCR-based methods known in the art. For example, the XL-PCR kit (Applied Biosystems), nested primers, and commercially available cDNA or genomic DNA libraries may be used to extend the nucleic acid sequence. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO primer analysis software (Molecular Biology Insights, Cascade CO) to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to a target molecule at temperatures from about 55C to about 68C. When extending a sequence to recover regulatory elements, it is preferable to use genomic, rather than cDNA libraries.

Hybridization

The cDNA and fragments thereof can be used in hybridization technologies for various purposes. A probe may be designed or derived from unique regions such as the 5' regulatory region or from a nonconserved region (i.e., 5' or 3' of the nucleotides encoding the conserved catalytic domain of the protein) and used in protocols to identify naturally occurring molecules encoding the STEAPRP, allelic variants, or related molecules. The probe may be DNA or RNA, may be single-stranded, and should have at least 50% sequence identity to any of the nucleic acid sequences, SEQ ID NOs:2-10. Hybridization probes may be produced using oligolabeling, nick translation, end-labeling, or PCR amplification in the presence of a reporter molecule. A vector containing the cDNA or a fragment thereof may be used to produce an mRNA probe in vitro by addition of an RNA polymerase and labeled nucleotides. These procedures may be conducted using commercially available kits such as those provided by APB.

The stringency of hybridization is determined by G+C content of the probe, salt concentration,

and temperature. In particular, stringency can be increased by reducing the concentration of salt or raising the hybridization temperature. Hybridization can be performed at low stringency with buffers, such as 5xSSC with 1% sodium dodecyl sulfate (SDS) at 60C, which permits the formation of a hybridization complex between nucleic acid sequences that contain some mismatches. Subsequent 5 washes are performed at higher stringency with buffers such as 0.2xSSC with 0.1% SDS at either 45C (medium stringency) or 68C (high stringency). At high stringency, hybridization complexes will remain stable only where the nucleic acids are completely complementary. In some membrane-based hybridizations, preferably 35% or most preferably 50%, formamide can be added to the hybridization solution to reduce the temperature at which hybridization is performed, and background signals can be 10 reduced by the use of detergents such as Sarkosyl or TRITON X-100 (Sigma-Aldrich, St. Louis MO) and a blocking agent such as denatured salmon sperm DNA. Selection of components and conditions for hybridization are well known to those skilled in the art and are reviewed in Ausubel (*supra*) and Sambrook *et al.* (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY.

15 Arrays may be prepared and analyzed using methods well known in the art. Oligonucleotides or cDNAs may be used as hybridization probes or targets to monitor the expression level of large numbers of genes simultaneously or to identify genetic variants, mutations, and single nucleotide polymorphisms. Arrays may be used to determine gene function; to understand the genetic basis of a condition, disease, or disorder; to diagnose a condition, disease, or disorder; and to develop and 20 monitor the activities of therapeutic agents. (See, e.g., Brennan *et al.* (1995) USPN 5,474,796; Schena *et al.* (1996) Proc Natl Acad Sci 93:10614-10619; Heller *et al.* (1997) Proc Natl Acad Sci 94:2150-2155; and Heller *et al.* (1997) USPN 5,605,662.)

Hybridization probes are also useful in mapping the naturally occurring genomic sequence. The probes may be hybridized to a particular chromosome, a specific region of a chromosome, or an 25 artificial chromosome construction. Such constructions include human artificial chromosomes (HAC), yeast artificial chromosomes (YAC), bacterial artificial chromosomes (BAC), bacterial P1 constructions, or the cDNAs of libraries made from single chromosomes.

Expression

Any one of a multitude of cDNAs encoding STEAPRP may be cloned into a vector and used 30 to express the protein, or portions thereof, in host cells. The nucleic acid sequence can be engineered by such methods as DNA shuffling (USPN 5,830,721) and site-directed mutagenesis to create new restriction sites, alter glycosylation patterns, change codon preference to increase expression in a particular host, produce splice variants, extend half-life, and the like. The expression vector may contain transcriptional and translational control elements (promoters, enhancers, specific initiation 35 signals, and polyadenylated 3' sequence) from various sources which have been selected for their

efficiency in a particular host. The vector, cDNA, and regulatory elements are combined using in vitro recombinant DNA techniques, synthetic techniques, and/or in vivo genetic recombination techniques well known in the art and described in Sambrook (supra, ch. 4, 8, 16 and 17).

A variety of host systems may be transformed with an expression vector. These include, but 5 are not limited to, bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems transformed with baculovirus expression vectors; plant cell systems transformed with expression vectors containing viral and/or bacterial elements, or animal cell systems (Ausubel supra, unit 16). For example, an adenovirus transcription/translation complex may be utilized in mammalian cells. After 10 sequences are ligated into the E1 or E3 region of the viral genome, the infective virus is used to transform and express the protein in host cells. The Rous sarcoma virus enhancer or SV40 or EBV-based vectors may also be used for high-level protein expression.

Routine cloning, subcloning, and propagation of nucleic acid sequences can be achieved using the multifunctional PBLUESCRIPT vector (Stratagene, La Jolla CA) or PSORT1 plasmid (Life 15 Technologies). Introduction of a nucleic acid sequence into the multiple cloning site of these vectors disrupts the lacZ gene and allows colorimetric screening for transformed bacteria. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence.

For long term production of recombinant proteins, the vector can be stably transformed into 20 cell lines along with a selectable or visible marker gene on the same or on a separate vector. After transformation, cells are allowed to grow for about 1 to 2 days in enriched media and then are transferred to selective media. Selectable markers, antimetabolite, antibiotic, or herbicide resistance genes, confer resistance to the relevant selective agent and allow growth and recovery of cells which successfully express the introduced sequences. Resistant clones identified either by survival on 25 selective media or by the expression of visible markers may be propagated using culture techniques. Visible markers are also used to estimate the amount of protein expressed by the introduced genes. Verification that the host cell contains the desired cDNA is based on DNA-DNA or DNA-RNA hybridizations or PCR amplification techniques.

The host cell may be chosen for its ability to modify a recombinant protein in a desired 30 fashion. Such modifications include acetylation, carboxylation, glycosylation, phosphorylation, lipidation, acylation and the like. Post-translational processing which cleaves a "prepro" form may also be used to specify protein targeting, folding, and/or activity. Different host cells available from the ATCC (Manassas VA) which have specific cellular machinery and characteristic mechanisms for post-translational activities may be chosen to ensure the correct modification and processing of the 35 recombinant protein.

Recovery of Proteins from Cell Culture

Heterologous moieties engineered into a vector for ease of purification include glutathione S-transferase (GST), 6xHis, FLAG, MYC, and the like. GST and 6-His are purified using commercially available affinity matrices such as immobilized glutathione and metal-chelate resins, respectively.

5 FLAG and MYC are purified using commercially available monoclonal and polyclonal antibodies.

For ease of separation following purification, a sequence encoding a proteolytic cleavage site may be part of the vector located between the protein and the heterologous moiety. Methods for recombinant protein expression and purification are discussed in Ausubel (*supra*, unit 16) and are commercially available.

10 Chemical Synthesis of Peptides

Proteins or portions thereof may be produced not only by recombinant methods, but also by using chemical methods well known in the art. Solid phase peptide synthesis may be carried out in a batchwise or continuous flow process which sequentially adds α -amino- and side chain-protected amino acid residues to an insoluble polymeric support via a linker group. A linker group such as 15 methylamine-derivatized polyethylene glycol is attached to poly(styrene-co-divinylbenzene) to form the support resin. The amino acid residues are N- α -protected by acid labile Boc (t-butyloxycarbonyl) or base-labile Fmoc (9-fluorenylmethoxycarbonyl). The carboxyl group of the protected amino acid is coupled to the amine of the linker group to anchor the residue to the solid phase support resin.

Trifluoroacetic acid or piperidine are used to remove the protecting group in the case of Boc or Fmoc, 20 respectively. Each additional amino acid is added to the anchored residue using a coupling agent or pre-activated amino acid derivative, and the resin is washed. The full length peptide is synthesized by sequential deprotection, coupling of derivitized amino acids, and washing with dichloromethane and/or N, N-dimethylformamide. The peptide is cleaved between the peptide carboxy terminus and the linker group to yield a peptide acid or amide. (Novabiochem 1997/98 Catalog and Peptide 25 Synthesis Handbook, San Diego CA pp. S1-S20). Automated synthesis may also be carried out on machines such as the ABI 431A peptide synthesizer (Applied Biosystems). A protein or portion thereof may be substantially purified by preparative high performance liquid chromatography and its composition confirmed by amino acid analysis or by sequencing (Creighton (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY).

30 Preparation and Screening of Antibodies

Various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with STEAPRP or any portion thereof. Adjuvants such as Freund's, mineral gels, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemacyanin (KLH), and dinitrophenol may be used to increase immunological 35 response. The oligopeptide, peptide, or portion of protein used to induce antibodies should consist of

at least about five amino acids, more preferably ten amino acids, which are identical to a portion of the natural protein. Oligopeptides may be fused with proteins such as KLH in order to produce antibodies to the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the 5 production of antibodies by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler *et al.* (1975) *Nature* 256:495-497; Kozbor *et al.* (1985) *J. Immunol Methods* 81:31-42; Cote *et al.* (1983) *Proc Natl Acad Sci* 80:2026-2030; and Cole *et al.* (1984) *Mol Cell Biol* 62:109-120.)

10 Alternatively, techniques described for antibody production may be adapted, using methods known in the art, to produce epitope-specific, single chain antibodies. Antibody fragments which contain specific binding sites for epitopes of the protein may also be generated. For example, such fragments include, but are not limited to, F(ab')2 fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 15 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse *et al.* (1989) *Science* 246:1275-1281.)

The STEAPRP or a portion thereof may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having the desired specificity. Numerous 20 protocols for competitive binding or immunoassays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Pound (1998) 25 Immunochemical Protocols, Humana Press, Totowa NJ).

Labeling of Molecules for Assay

A wide variety of reporter molecules and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid, amino acid, and antibody assays. Synthesis of labeled molecules may be achieved using commercially available kits (Promega, Madison WI) for 30 incorporation of a labeled nucleotide such as ³²P-dCTP (APB), Cy3-dCTP or Cy5-dCTP (Operon Technologies, Alameda CA), or amino acid such as ³⁵S-methionine (APB). Nucleotides and amino acids may be directly labeled with a variety of substances including fluorescent, chemiluminescent, or chromogenic agents, and the like, by chemical conjugation to amines, thiols and other groups present in the molecules using reagents such as BIODIPY or FITC (Molecular Probes, Eugene OR).

35 DIAGNOSTICS

The cDNAs, fragments, oligonucleotides, complementary RNA and DNA molecules, and PNAs and may be used to detect and quantify differential gene expression for diagnosis of a disorder. Similarly antibodies which specifically bind STEAPRP may be used to quantitate the protein. Disorders associated with differential expression include prostate cell proliferative disorders, 5 particularly prostate hyperplasia and prostate cancer. The diagnostic assay may use hybridization or amplification technology to compare gene expression in a biological sample from a patient to standard samples in order to detect differential gene expression. Qualitative or quantitative methods for this comparison are well known in the art.

For example, the cDNA or probe may be labeled by standard methods and added to a 10 biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes, is quantified and compared with a standard value. If complex formation in the patient sample is significantly altered (higher or lower) in comparison to either a normal or disease standard, then differential expression indicates the presence of a disorder.

15 In order to provide standards for establishing differential expression, normal and disease expression profiles are established. This is accomplished by combining a sample taken from normal subjects, either animal or human, with a cDNA under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used. Standard 20 values obtained in this manner may be compared with values obtained from samples from patients who were diagnosed with a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular disorder is used to diagnose that disorder.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies or in clinical trials or to monitor the treatment of an individual patient. 25 Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

Immunological Methods

30 Detection and quantification of a protein using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed. (See, e.g., 35 Coligan *et al.* (1997) Current Protocols in Immunology, Wiley-Interscience, New York NY; and

Pound, supra.)

THERAPEUTICS

Chemical and structural similarity, in particular the six transmembrane domains, exists between regions of STEAPRP (SEQ ID NO:1) and human STEAP (g6572948; SEQ ID NO:11) as shown in Figures 2A, 2B, and 2C. In addition, differential expression is highly associated with LNCaP prostate carcinoma cells and prostate tissues and with prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer as shown in Tables 1-3. STEAPRP clearly plays a role in prostate cell proliferative disorders, particularly prostate hyperplasia and prostate cancer.

In the treatment of conditions associated with increased expression of the STEAPRP, it is 10 desirable to decrease expression or protein activity. In one embodiment, the an inhibitor, antagonist, or antibody of the protein may be administered to a subject to treat a condition associated with increased expression or activity. In another embodiment, a pharmaceutical composition comprising an inhibitor, antagonist or antibody in conjunction with a pharmaceutical carrier may be administered to a subject to treat a condition associated with the increased expression or activity of the endogenous 15 protein. In an additional embodiment, a vector expressing the complement of the cDNA or fragments thereof may be administered to a subject to treat the disorder.

In the treatment of conditions associated with decreased expression of the STEAPRP, it is desirable to increase expression or protein activity. In one embodiment, the protein, an agonist, or enhancer may be administered to a subject to treat a condition associated with decreased expression or 20 activity. In another embodiment, a pharmaceutical composition comprising the protein, an agonist or enhancer in conjunction with a pharmaceutical carrier may be administered to a subject to treat a condition associated with the decreased expression or activity of the endogenous protein. In an additional embodiment, a vector expressing cDNA may be administered to a subject to treat the disorder.

25 Any of the cDNAs, complementary molecules, or fragments thereof, proteins or portions thereof, vectors delivering these nucleic acid molecules or expressing the proteins, and their ligands may be administered in combination with other therapeutic agents. Selection of the agents for use in combination therapy may be made by one of ordinary skill in the art according to conventional pharmaceutical principles. A combination of therapeutic agents may act synergistically to affect 30 treatment of a particular disorder at a lower dosage of each agent.

Modification of Gene Expression Using Nucleic Acids

Gene expression may be modified by designing complementary or antisense molecules (DNA, RNA, or PNA) to the control, 5', 3', or other regulatory regions of the gene encoding STEAPRP. Oligonucleotides designed to inhibit transcription initiation are preferred. Similarly, inhibition can be 35 achieved using triple helix base-pairing which inhibits the binding of polymerases, transcription

factors, or regulatory molecules (Gee *et al.* In: Huber and Carr (1994) Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary molecule may also be designed to block translation by preventing binding between ribosomes and mRNA. In one alternative, a library or plurality of cDNAs may be screened to identify those which specifically bind a 5 regulatory, nontranslated sequence.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA followed by endonucleolytic cleavage at sites such as GUA, GUU, and GUC. Once such sites are identified, an oligonucleotide with the same sequence may be 10 evaluated for secondary structural features which would render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing their hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary nucleic acids and ribozymes of the invention may be prepared via recombinant expression, *in vitro* or *in vivo*, or using solid phase phosphoramidite chemical synthesis. 15 In addition, RNA molecules may be modified to increase intracellular stability and half-life by addition of flanking sequences at the 5' and/or 3' ends of the molecule or by the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages within the backbone of the molecule. Modification is inherent in the production of PNAs and can be extended to other nucleic acid molecules. Either the inclusion of nontraditional bases such as inosine, queosine, and 20 wybutosine, and/or the modification of adenine, cytidine, guanine, thymine, and uridine with acetyl-, methyl-, thio- groups renders the molecule less available to endogenous endonucleases.

Screening and Purification Assays

The cDNA encoding STEAPRP may be used to screen a library of molecules or compounds for specific binding affinity. The libraries may be aptamers, DNA molecules, RNA molecules, PNAs, 25 peptides, proteins such as transcription factors, enhancers, repressors, and other ligands which regulate the activity, replication, transcription, or translation of the endogenous gene. The assay involves combining a polynucleotide with a library of molecules under conditions allowing specific binding, and detecting specific binding to identify at least one molecule which specifically binds the single-stranded or double-stranded molecule.

30 In one embodiment, the cDNA of the invention may be incubated with a plurality of purified molecules or compounds and binding activity determined by methods well known in the art, e.g., a gel-retardation assay (USPN 6,010,849) or a reticulocyte lysate transcriptional assay. In another embodiment, the cDNA may be incubated with nuclear extracts from biopsied and/or cultured cells and tissues. Specific binding between the cDNA and a molecule or compound in the nuclear extract is 35 initially determined by gel shift assay and may be later confirmed by recovering and raising antibodies

against that molecule or compound. When these antibodies are added into the assay, they cause a supershift in the gel-retardation assay.

In another embodiment, the cDNA may be used to purify a molecule or compound using affinity chromatography methods well known in the art. In one embodiment, the cDNA is chemically 5 reacted with cyanogen bromide groups on a polymeric resin or gel. Then a sample is passed over and reacts with or binds to the cDNA. The molecule or compound which is bound to the cDNA may be released from the cDNA by increasing the salt concentration of the flow-through medium and collected.

In a further embodiment, the protein or a portion thereof may be used to purify a ligand from a 10 sample. A method for using a protein or a portion thereof to purify a ligand would involve combining the protein or a portion thereof with a sample under conditions to allow specific binding, detecting specific binding between the protein and ligand, recovering the bound protein, and using an appropriate chaotropic agent to separate the protein from the purified ligand.

In a preferred embodiment, STEAPRP may be used to screen a plurality of molecules or 15 compounds in any of a variety of screening assays. The portion of the protein employed in such screening may be free in solution, affixed to an abiotic or biotic substrate (e.g. borne on a cell surface), or located intracellularly. For example, in one method, viable or fixed prokaryotic host cells that are stably transformed with recombinant nucleic acids that have expressed and positioned a peptide on their cell surface can be used in screening assays. The cells are screened against a plurality 20 or libraries of ligands, and the specificity of binding or formation of complexes between the expressed protein and the ligand may be measured. Specific binding between the protein and molecule may be measured. Depending on the particular kind of library being screened, the assay may be used to identify DNA molecules, RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs or any other ligand, which 25 specifically binds the protein.

In one aspect, this invention contemplates a method for high throughput screening using very small assay volumes and very small amounts of test compound as described in USPN 5,876,946, incorporated herein by reference. This method is used to screen large numbers of molecules and compounds via specific binding. In another aspect, this invention also contemplates the use of 30 competitive drug screening assays in which neutralizing antibodies capable of binding the protein specifically compete with a test compound capable of binding to the protein. Molecules or compounds identified by screening may be used in a mammalian model system to evaluate their toxicity, diagnostic, or therapeutic potential.

Pharmacology

35 Pharmaceutical compositions are those substances wherein the active ingredients are

contained in an effective amount to achieve a desired and intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose may be estimated initially either in cell culture assays or in animal models. The animal model is also used to achieve a desirable concentration range and route of administration. Such information may then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or inhibitor which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity of such agents may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED₅₀ (the dose 10 therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it may be expressed as the ratio, LD₅₀/ED₅₀. Pharmaceutical compositions which exhibit large therapeutic indexes are preferred. The data obtained from cell culture assays and animal studies are used in formulating a range of dosage for human use.

15 **Model Systems**

Animal models may be used as bioassays where they exhibit a phenotypic response similar to that of humans and where exposure conditions are relevant to human exposures. Mammals are the most common models, and most infectious agent, cancer, drug, and toxicity studies are performed on rodents such as rats or mice because of low cost, availability, lifespan, reproductive potential, and 20 abundant reference literature. Inbred and outbred rodent strains provide a convenient model for investigation of the physiological consequences of under- or over-expression of genes of interest and for the development of methods for diagnosis and treatment of diseases. A mammal inbred to over-express a particular gene (for example, secreted in milk) may also serve as a convenient source of the protein expressed by that gene.

25 **Toxicology**

Toxicology is the study of the effects of agents on living systems. The majority of toxicity studies are performed on rats or mice. Observation of qualitative and quantitative changes in physiology, behavior, homeostatic processes, and lethality in the rats or mice are used to generate a toxicity profile and to assess potential consequences on human health following exposure to the agent. 30 Genetic toxicology identifies and analyzes the effect of an agent on the rate of endogenous, spontaneous, and induced genetic mutations. Genotoxic agents usually have common chemical or physical properties that facilitate interaction with nucleic acids and are most harmful when chromosomal aberrations are transmitted to progeny. Toxicological studies may identify agents that increase the frequency of structural or functional abnormalities in the tissues of the progeny if 35 administered to either parent before conception, to the mother during pregnancy, or to the developing

organism. Mice and rats are most frequently used in these tests because their short reproductive cycle allows the production of the numbers of organisms needed to satisfy statistical requirements.

Acute toxicity tests are based on a single administration of an agent to the subject to determine the symptomology or lethality of the agent. Three experiments are conducted: 1) an initial 5 dose-range-finding experiment, 2) an experiment to narrow the range of effective doses, and 3) a final experiment for establishing the dose-response curve.

Subchronic toxicity tests are based on the repeated administration of an agent. Rat and dog are commonly used in these studies to provide data from species in different families. With the exception of carcinogenesis, there is considerable evidence that daily administration of an agent at 10 high-dose concentrations for periods of three to four months will reveal most forms of toxicity in adult animals.

Chronic toxicity tests, with a duration of a year or more, are used to demonstrate either the absence of toxicity or the carcinogenic potential of an agent. When studies are conducted on rats, a minimum of three test groups plus one control group are used, and animals are examined and 15 monitored at the outset and at intervals throughout the experiment.

Transgenic Animal Models

Transgenic rodents that over-express or under-express a gene of interest may be inbred and used to model human diseases or to test therapeutic or toxic agents. (See, e.g., USPN 5,175,383 and USPN 5,767,337.) In some cases, the introduced gene may be activated at a specific time in a specific 20 tissue type during fetal or postnatal development. Expression of the transgene is monitored by analysis of phenotype, of tissue-specific mRNA expression, or of serum and tissue protein levels in transgenic animals before, during, and after challenge with experimental drug therapies.

Embryonic Stem Cells

Embryonic (ES) stem cells isolated from rodent embryos retain the potential to form 25 embryonic tissues. When ES cells are placed inside a carrier embryo, they resume normal development and contribute to tissues of the live-born animal. ES cells are the preferred cells used in the creation of experimental knockout and knockin rodent strains. Mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and are grown under culture conditions well known in the art. Vectors used to produce a transgenic strain contain a disease gene candidate 30 and a marker gene, the latter serves to identify the presence of the introduced disease gene. The vector is transformed into ES cells by methods well known in the art, and transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

35 ES cells derived from human blastocysts may be manipulated in vitro to differentiate into at

least eight separate cell lineages. These lineages are used to study the differentiation of various cell types and tissues in vitro, and they include endoderm, mesoderm, and ectodermal cell types which differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes.

Knockout Analysis

5 In gene knockout analysis, a region of a mammalian gene is enzymatically modified to include a non-mammalian gene such as the neomycin phosphotransferase gene (neo; Capecchi (1989) Science 244:1288-1292). The modified gene is transformed into cultured ES cells and integrates into the endogenous genome by homologous recombination. The inserted sequence disrupts transcription and translation of the endogenous gene. Transformed cells are injected into rodent blastulae, and the
10 blastulae are implanted into pseudopregnant dams. Transgenic progeny are crossbred to obtain homozygous inbred lines which lack a functional copy of the mammalian gene. In one example, the mammalian gene is a human gene.

Knockin Analysis

ES cells can be used to create knockin humanized animals (pigs) or transgenic animal models
15 (mice or rats) of human diseases. With knockin technology, a region of a human gene is injected into animal ES cells, and the human sequence integrates into the animal cell genome. Transformed cells are injected into blastulae and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of the analogous human condition. These methods have been used to model several human
20 diseases.

Non-Human Primate Model

The field of animal testing deals with data and methodology from basic sciences such as physiology, genetics, chemistry, pharmacology and statistics. These data are paramount in evaluating the effects of therapeutic agents on non-human primates as they can be related to human health.
25 Monkeys are used as human surrogates in vaccine and drug evaluations, and their responses are relevant to human exposures under similar conditions. Cynomolgus and Rhesus monkeys (Macaca fascicularis and Macaca mulatta, respectively) and Common Marmosets (Callithrix jacchus) are the most common non-human primates (NHPs) used in these investigations. Since great cost is associated with developing and maintaining a colony of NHPs, early research and toxicological studies are
30 usually carried out in rodent models. In studies using behavioral measures such as drug addiction, NHPs are the first choice test animal. In addition, NHPs and individual humans exhibit differential sensitivities to many drugs and toxins and can be classified as a range of phenotypes from "extensive metabolizers" to "poor metabolizers" of these agents.

In additional embodiments, the cDNAs which encode the protein may be used in any
35 molecular biology techniques that have yet to be developed, provided the new techniques rely on

properties of cDNAs that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

EXAMPLES

5 The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention. The preparation of the human brain dentate nucleus (BRAWTDR02) library will be described.

I cDNA Library Construction

The BRAWTDR02 cDNA library was constructed from brain dentate nucleus tissue removed 10 from a 55-year-old Caucasian female (specimen #A98-58) who died from cholangiocarcinoma. The frozen tissue was homogenized and lysed in TRIZOL reagent (0.8 g tissue/12 ml; Life Technologies) using a POLYTRON homogenizer (Brinkmann Instruments, Westbury NJ). The lysate was centrifuged over a 5.7 M CsCl cushion using an SW28 rotor in an L8-70M ultracentrifuge (Beckman Coulter, Fullerton CA) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted 15 with acid phenol, pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and treated with DNase at 37C. The RNA was reextracted and precipitated as before. The mRNA was isolated with the OLIGOTEX kit (Qiagen, Chatsworth CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERSCRIPT 20 plasmid system (Life Technologies) which contains a NotI primer-adaptor designed to prime the first strand cDNA synthesis at the poly(A) tail of mRNAs. Double stranded cDNA was blunted, ligated to EcoRI adaptors and digested with NotI (New England Biolabs, Beverly MA). The cDNAs were fractionated on a SEPHAROSE CL4B column (APB), and those cDNAs exceeding 400 bp were ligated into pcDNA2.1 plasmid (Invitrogen, Carlsbad CA). The plasmid pcDNA2.1 was subsequently 25 transformed into DH5 α competent cells (Life Technologies).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using either the MINIPREP kit (Edge Biosystems, Gaithersburg MD) or the REAL PREP 96 plasmid kit (Qiagen). A kit consists of a 96-well block with reagents for 960 purifications. The recommended protocol was employed except for 30 the following changes: 1) the bacteria were cultured in 1 ml of sterile TERRIFIC BROTH (APB) with carbenicillin at 25 mg/l and glycerol at 0.4%; 2) after inoculation, the cells were cultured for 19 hours and then lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4C.

35 The cDNAs were prepared for sequencing using the MICROLAB 2200 system (Hamilton) in

combination with the DNA ENGINE thermal cyclers (MJ Research). The cDNAs were sequenced by the method of Sanger and Coulson (1975; J Mol Biol 94:441-448) using an ABI PRISM 377 sequencing system (Applied Biosystems) or the MEGABACE 1000 DNA sequencing system (APB). Most of the isolates were sequenced according to standard ABI protocols and kits (Applied Biosystems) with solution volumes of 0.25x-1.0x concentrations. In the alternative, cDNAs were sequenced using solutions and dyes from APB.

III Extension of cDNA Sequences

The cDNAs were extended using the cDNA clone and oligonucleotide primers. One primer was synthesized to initiate 5' extension of the known fragment, and the other, to initiate 3' extension of the known fragment. The initial primers were designed using commercially available primer analysis software to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68C to about 72C. Any stretch of nucleotides that would result in hairpin structures and primer-primer dimerizations was avoided.

Selected cDNA libraries were used as templates to extend the sequence. If more than one extension was necessary, additional or nested sets of primers were designed. Preferred libraries have been size-selected to include larger cDNAs and random primed to contain more sequences with 5' or upstream regions of genes. Genomic libraries are used to obtain regulatory elements, especially extension into the 5' promoter binding region.

High fidelity amplification was obtained by PCR using methods such as that taught in USPN 5,932,451. PCR was performed in 96-well plates using the DNA ENGINE thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and β-mercaptoethanol, Taq DNA polymerase (APB), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B (Incyte Genomics): Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C. In the alternative, the parameters for primer pair T7 and SK+ (Stratagene) were as follows: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 57C, one min; Step 4: 68C, two min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68C, five min; Step 7: storage at 4C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% reagent in 1x TE, v/v; Molecular Probes) and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning, Acton MA) and allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1% agarose minigel to determine which reactions were successful in extending the sequence.

The extended clones were desalted, concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC18 vector (APB). For shotgun sequences, the digested nucleotide sequences were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, 5 and the agar was digested with AGARACE enzyme (Promega). Extended clones were religated using T4 DNA ligase (New England Biolabs) into pUC18 vector (APB), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into E. coli competent cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37C in 384-well plates in LB/2x carbenicillin liquid media.

10 The cells were lysed, and DNA was amplified using primers, Taq DNA polymerase (APB) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94C, three min; Step 2: 94C, 15 sec; Step 3: 60C, one min; Step 4: 72C, two min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72C, five min; Step 7: storage at 4C. DNA was quantified using PICOGREEN quantitation reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified 15 using the conditions described above. Samples were diluted with 20% dimethylsulfoxide (DMSO; 1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT cycle sequencing kit (APB) or the ABI PRISM BIGDYE terminator cycle sequencing kit (Applied Biosystems).

IV Homology Searching of cDNA Clones and Their Deduced Proteins

20 The cDNAs of the Sequence Listing or their deduced amino acid sequences were used to query databases such as GenBank, SwissProt, BLOCKS, and the like. These databases that contain previously identified and annotated sequences or domains were searched using BLAST or BLAST2 to produce alignments and to determine which sequences were exact matches or homologs. The alignments were to sequences of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. 25 Alternatively, algorithms such as the one described in Smith and Smith (1992, Protein Engineering 5:35-51) could have been used to deal with primary sequence patterns and secondary structure gap penalties. All of the sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

As detailed in Karlin (*supra*), BLAST matches between a query sequence and a database 30 sequence were evaluated statistically and only reported when they satisfied the threshold of 10^{-25} for nucleotides and 10^{-14} for peptides. Homology was also evaluated by product score calculated as follows: the % nucleotide or amino acid identity [between the query and reference sequences] in BLAST is multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences] and then divided by 100. In comparison with hybridization procedures used in 35 the laboratory, the stringency for an exact match was set from a lower limit of about 40 (with 1-2%

error due to uncalled bases) to a 100% match of about 70.

The BLAST software suite (NCBI, Bethesda MD;

<http://www.ncbi.nlm.nih.gov/gorf/bl2.html>), includes various sequence analysis programs including "blastn" that is used to align nucleotide sequences and BLAST2 that is used for direct pairwise comparison of either nucleotide or amino acid sequences. BLAST programs are commonly used with gap and other parameters set to default settings, e.g.: Matrix: BLOSUM62; Reward for match: 1; Penalty for mismatch: -2; Open Gap: 5 and Extension Gap: 2 penalties; Gap x drop-off: 50; Expect: 10; Word Size: 11; and Filter: on. Identity is measured over the entire length of a sequence. Brenner et al. (1998; Proc Natl Acad Sci 95:6073-6078, incorporated herein by reference) analyzed BLAST for its ability to identify structural homologs by sequence identity and found 30% identity is a reliable threshold for sequence alignments of at least 150 residues and 40%, for alignments of at least 70 residues.

The cDNAs of this application were compared with assembled consensus sequences or templates found in the LIFESEQ GOLD database (Incyte Genomics). Component sequences from cDNA, extension, full length, and shotgun sequencing projects were subjected to PHRED analysis and assigned a quality score. All sequences with an acceptable quality score were subjected to various pre-processing and editing pathways to remove low quality 3' ends, vector and linker sequences, polyA tails, Alu repeats, mitochondrial and ribosomal sequences, and bacterial contamination sequences. Edited sequences had to be at least 50 bp in length, and low-information sequences and repetitive elements such as dinucleotide repeats, Alu repeats, and the like, were replaced by "Ns" or masked.

Edited sequences were subjected to assembly procedures in which the sequences were assigned to gene bins. Each sequence could only belong to one bin, and sequences in each bin were assembled to produce a template. Newly sequenced components were added to existing bins using BLAST and CROSSMATCH. To be added to a bin, the component sequences had to have a BLAST quality score greater than or equal to 150 and an alignment of at least 82% local identity. The sequences in each bin were assembled using PHRAP. Bins with several overlapping component sequences were assembled using DEEP PHRAP. The orientation of each template was determined based on the number and orientation of its component sequences.

Bins were compared to one another, and those having local similarity of at least 82% were combined and reassembled. Bins having templates with less than 95% local identity were split. Templates were subjected to analysis by STITCHER/EXON MAPPER algorithms that determine the probabilities of the presence of splice variants, alternatively spliced exons, splice junctions, differential expression of alternative spliced genes across tissue types or disease states, and the like. Assembly procedures were repeated periodically, and templates were annotated using BLAST against

GenBank databases such as GBpri. An exact match was defined as having from 95% local identity over 200 base pairs through 100% local identity over 100 base pairs and a homolog match as having an E-value (or probability score) of $\leq 1 \times 10^{-8}$. The templates were also subjected to frameshift FASTx against GENPEPT, and homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. Template 5 analysis and assembly was described in USSN 09/276,534, filed March 25, 1999.

Following assembly, templates were subjected to BLAST, motif, and other functional analyses and categorized in protein hierarchies using methods described in USSN 08/812,290 and USSN 08/811,758, both filed March 6, 1997; in USSN 08/947,845, filed October 9, 1997; and in USSN 09/034,807, filed March 4, 1998. Then templates were analyzed by translating each template 10 in all three forward reading frames and searching each translation against the PFAM database of hidden Markov model-based protein families and domains using the HMMER software package (Washington University School of Medicine, St. Louis MO; <http://pfam.wustl.edu/>). The cDNA was further analyzed using MACDNASIS PRO software (Hitachi Software Engineering), and LASERGENE software (DNASTAR) and queried against public databases such as the GenBank 15 rodent, mammalian, vertebrate, prokaryote, and eukaryote databases, SwissProt, BLOCKS, PRINTS, PFAM, and Prosite.

V Chromosome Mapping

Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and 20 Généthon are used to determine if any of the cDNAs presented in the Sequence Listing have been mapped. Any of the fragments of the cDNA encoding STEAPRP that have been mapped result in the assignment of all related regulatory and coding sequences mapping to the same location. The genetic map locations are described as ranges, or intervals, of human chromosomes. The map position of an interval, in cM (which is roughly equivalent to 1 megabase of human DNA), is measured relative to 25 the terminus of the chromosomal p-arm.

VI Hybridization Technologies and Analyses

Immobilization of cDNAs on a Substrate

The cDNAs are applied to a substrate by one of the following methods. A mixture of cDNAs is fractionated by gel electrophoresis and transferred to a nylon membrane by capillary transfer. 30 Alternatively, the cDNAs are individually ligated to a vector and inserted into bacterial host cells to form a library. The cDNAs are then arranged on a substrate by one of the following methods. In the first method, bacterial cells containing individual clones are robotically picked and arranged on a nylon membrane. The membrane is placed on LB agar containing selective agent (carbenicillin, kanamycin, ampicillin, or chloramphenicol depending on the vector used) and incubated at 37C for 16 35 hr. The membrane is removed from the agar and consecutively placed colony side up in 10% SDS,

denaturing solution (1.5 M NaCl, 0.5 M NaOH), neutralizing solution (1.5 M NaCl, 1 M Tris, pH 8.0), and twice in 2xSSC for 10 min each. The membrane is then UV irradiated in a STRATALINKER UV-crosslinker (Stratagene).

In the second method, cDNAs are amplified from bacterial vectors by thirty cycles of PCR 5 using primers complementary to vector sequences flanking the insert. PCR amplification increases a starting concentration of 1-2 ng nucleic acid to a final quantity greater than 5 μ g. Amplified nucleic acids from about 400 bp to about 5000 bp in length are purified using SEPHACRYL-400 beads (APB). Purified nucleic acids are arranged on a nylon membrane manually or using a dot/slot blotting manifold and suction device and are immobilized by denaturation, neutralization, and UV irradiation 10 as described above. Purified nucleic acids are robotically arranged and immobilized on polymer-coated glass slides using the procedure described in USPN 5,807,522. Polymer-coated slides are prepared by cleaning glass microscope slides (Corning, Acton MA) by ultrasound in 0.1% SDS and acetone, etching in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), coating with 0.05% aminopropyl silane (Sigma Aldrich) in 95% ethanol, and curing in a 110C oven. The slides are 15 washed extensively with distilled water between and after treatments. The nucleic acids are arranged on the slide and then immobilized by exposing the array to UV irradiation using a STRATALINKER UV-crosslinker (Stratagene). Arrays are then washed at room temperature in 0.2% SDS and rinsed three times in distilled water. Non-specific binding sites are blocked by incubation of arrays in 0.2% casein in phosphate buffered saline (PBS; Tropix, Bedford MA) for 30 min at 60C; then the arrays are 20 washed in 0.2% SDS and rinsed in distilled water as before.

Probe Preparation for Membrane Hybridization

Hybridization probes derived from the cDNAs of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA in membrane-based hybridizations. Probes are prepared by diluting the cDNAs to a concentration of 40-50 ng in 45 μ l TE buffer, denaturing by heating to 25 100C for five min, and briefly centrifuging. The denatured cDNA is then added to a REDIPRIME tube (APB); gently mixed until blue color is evenly distributed, and briefly centrifuged. Five μ l of [32 P]dCTP is added to the tube, and the contents are incubated at 37C for 10 min. The labeling reaction is stopped by adding 5 μ l of 0.2M EDTA, and probe is purified from unincorporated 30 nucleotides using a PROBEQUANT G-50 microcolumn (APB). The purified probe is heated to 100C for five min, snap cooled for two min on ice, and used in membrane-based hybridizations as described below.

Probe Preparation for Polymer Coated Slide Hybridization

Hybridization probes derived from mRNA isolated from samples are employed for screening cDNAs of the Sequence Listing in array-based hybridizations. Probe is prepared using the GEMbright 35 kit (Incyte Genomics) by diluting mRNA to a concentration of 200 ng in 9 μ l TE buffer and adding 5

μl 5x buffer, 1 μl 0.1 M DTT, 3 μl Cy3 or Cy5 labeling mix, 1 μl RNase inhibitor, 1 μl reverse transcriptase, and 5 μl 1x yeast control mRNAs. Yeast control mRNAs are synthesized by in vitro transcription from noncoding yeast genomic DNA (W. Lei, unpublished). As quantitative controls, one set of control mRNAs at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng are diluted into reverse transcription 5 reaction mixture at ratios of 1:100,000, 1:10,000, 1:1000, and 1:100 (w/w) to sample mRNA respectively. To examine mRNA differential expression patterns, a second set of control mRNAs are diluted into reverse transcription reaction mixture at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, and 25:1 (w/w). The reaction mixture is mixed and incubated at 37C for two hr. The reaction mixture is then incubated for 20 min at 85C, and probes are purified using two successive CHROMA SPIN+TE 30 10 columns (Clontech, Palo Alto CA). Purified probe is ethanol precipitated by diluting probe to 90 μl in DEPC-treated water, adding 2 μl 1mg/ml glycogen, 60 μl 5 M sodium acetate, and 300 μl 100% ethanol. The probe is centrifuged for 20 min at 20,800xg, and the pellet is resuspended in 12 μl resuspension buffer, heated to 65C for five min, and mixed thoroughly. The probe is heated and mixed as before and then stored on ice. Probe is used in high density array-based hybridizations as 15 described below.

Membrane-based Hybridization

Membranes are pre-hybridized in hybridization solution containing 1% Sarkosyl and 1x high phosphate buffer (0.5 M NaCl, 0.1 M Na₂HPO₄, 5 mM EDTA, pH 7) at 55C for two hr. The probe, diluted in 15 ml fresh hybridization solution, is then added to the membrane. The membrane is 20 hybridized with the probe at 55C for 16 hr. Following hybridization, the membrane is washed for 15 min at 25C in 1mM Tris (pH 8.0), 1% Sarkosyl, and four times for 15 min each at 25C in 1mM Tris (pH 8.0). To detect hybridization complexes, XOMAT-AR film (Eastman Kodak, Rochester NY) is exposed to the membrane overnight at -70C, developed, and examined visually.

Polymer Coated Slide-based Hybridization

25 Probe is heated to 65C for five min, centrifuged five min at 9400 rpm in a 5415C microcentrifuge (Eppendorf Scientific, Westbury NY), and then 18 μl is aliquoted onto the array surface and covered with a coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5xSSC in a corner of the chamber. The chamber containing the arrays is 30 incubated for about 6.5 hr at 60C. The arrays are washed for 10 min at 45C in 1xSSC, 0.1% SDS, and three times for 10 min each at 45C in 0.1xSSC, and dried.

Hybridization reactions are performed in absolute or differential hybridization formats. In the absolute hybridization format, probe from one sample is hybridized to array elements, and signals are detected after hybridization complexes form. Signal strength correlates with probe mRNA levels in 35 the sample. In the differential hybridization format, differential expression of a set of genes in two

biological samples is analyzed. Probes from the two samples are prepared and labeled with different labeling moieties. A mixture of the two labeled probes is hybridized to the array elements, and signals are examined under conditions in which the emissions from the two different labels are individually detectable. Elements on the array that are hybridized to substantially equal numbers of probes derived 5 from both biological samples give a distinct combined fluorescence (Shalon WO95/35505).

Hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Melville NY). The slide containing the array is 10 placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective with a resolution of 20 micrometers. In the differential hybridization format, the two fluorophores are sequentially excited by the laser. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are 15 used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. The sensitivity of the scans is calibrated using the signal intensity generated by the yeast control mRNAs added to the probe mix. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000.

20 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and 25 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using the emission spectrum for each fluorophore. A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal 30 analysis is the GEMTOOLS program (Incyte Genomics).

VII Electronic Analysis

BLAST was used to search for identical or related molecules in the GenBank or LIFESEQ databases (Incyte Genomics). The product score for human and rat sequences was calculated as follows: the BLAST score is multiplied by the % nucleotide identity and the product is divided by (5 35 times the length of the shorter of the two sequences), such that a 100% alignment over the length of

the shorter sequence gives a product score of 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and with a product score of at least 70, the match will be exact. Similar or related molecules are usually identified by selecting those 5 which show product scores between 8 and 40.

Electronic northern analysis was performed at a product score of 70 and is shown in Tables 1 and 2. All sequences and cDNA libraries in the LIFESEQ database were categorized by system, organ/tissue and cell type. The categories included cardiovascular system, connective tissue, digestive system, embryonic structures, endocrine system, exocrine glands, female and male genitalia, germ 10 cells, hemic/immune system, liver, musculoskeletal system, nervous system, pancreas, respiratory system, sense organs, skin, stomatognathic system, unclassified/mixed, and the urinary tract. For each category, the number of libraries in which the sequence was expressed were counted and shown over the total number of libraries in that category. In a non-normalized library, expression levels of two or more are significant.

15 VIII Complementary Molecules

Molecules complementary to the cDNA, from about 5 (PNA) to about 5000 bp (complement of a cDNA insert), are used to detect or inhibit gene expression. Detection is described in Example VI. To inhibit transcription by preventing promoter binding, the complementary molecule is designed to bind to the most unique 5' sequence and includes nucleotides of the 5' UTR upstream of the 20 initiation codon of the open reading frame. Complementary molecules include genomic sequences (such as enhancers or introns) and are used in "triple helix" base pairing to compromise the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. To inhibit translation, a complementary molecule is designed to prevent ribosomal binding to the mRNA encoding the protein.

25 Complementary molecules are placed in expression vectors and used to transform a cell line to test efficacy; into an organ, tumor, synovial cavity, or the vascular system for transient or short term therapy; or into a stem cell, zygote, or other reproducing lineage for long term or stable gene therapy. Transient expression lasts for a month or more with a non-replicating vector and for three months or more if appropriate elements for inducing vector replication are used in the transformation/expression 30 system.

Stable transformation of appropriate dividing cells with a vector encoding the complementary molecule produces a transgenic cell line, tissue, or organism (USPN 4,736,866). Those cells that assimilate and replicate sufficient quantities of the vector to allow stable integration also produce enough complementary molecules to compromise or entirely eliminate activity of the cDNA encoding 35 the protein.

IX Selection of Sequences, Microarray Preparation and Use

Incyte clones represent template sequences derived from the LIFESEQ GOLD assembled human sequence database (Incyte Genomics). In cases where more than one clone was available for a particular template, the 5'-most clone in the template was used on the microarray. The HUMAN 5 GENOME GEM series 1-3 microarrays (Incyte Genomics) contain 28,626 array elements which represent 10,068 annotated clusters and 18,558 unannotated clusters. For the UNIGEM series microarrays (Incyte Genomics), Incyte clones were mapped to non-redundant Unigene clusters (Unigene database (build 46), NCBI; Shuler (1997) J Mol Med 75:694-698), and the 5' clone with the strongest BLAST alignment (at least 90% identity and 100 bp overlap) was chosen, verified, and used 10 in the construction of the microarray. The UNIGEM V microarray (Incyte Genomics) contains 7075 array elements which represent 4610 annotated genes and 2,184 unannotated clusters.

To construct microarrays, cDNAs were amplified from bacterial cells using primers complementary to vector sequences flanking the cDNA insert. Thirty cycles of PCR increased the initial quantity of cDNAs from 1-2 ng to a final quantity of greater than 5 μ g. Amplified cDNAs were 15 then purified using SEPHACRYL-400 columns (APB). Purified cDNAs were immobilized on polymer-coated glass slides. Glass microscope slides (Corning, Corning NY) were cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides were etched in 4% hydrofluoric acid (VWR Scientific Products, West Chester PA), washed thoroughly in distilled water, and coated with 0.05% aminopropyl silane (Sigma 20 Aldrich) in 95% ethanol. Coated slides were cured in a 110°C oven. cDNAs were applied to the coated glass substrate using a procedure described in USPN 5,807,522. One microliter of the cDNA at an average concentration of 100 ng/ μ l was loaded into the open capillary printing element by a high-speed robotic apparatus which then deposited about 5 nl of cDNA per slide.

Microarrays were UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene), 25 and then washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites were blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (Tropix, Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

X Preparation of Samples

30 LNCaP (ATCC, Manassus VA) is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens, produce prostatic acid phosphatase, and express androgen receptors. Gene expression profiles of LNCaP prostate carcinoma cells were compared to those of nontumorigenic primary prostate epithelial PrEC cells.

35 XI Expression of STEAPRP

Expression and purification of the protein are achieved using either a mammalian cell expression system or an insect cell expression system. The pUB6/V5-His vector system (Invitrogen, Carlsbad CA) is used to express STEAPRP in CHO cells. The vector contains the selectable bsd gene, multiple cloning sites, the promoter/enhancer sequence from the human ubiquitin C gene, a C-terminal V5 epitope for antibody detection with anti-V5 antibodies, and a C-terminal polyhistidine (6xHis) sequence for rapid purification on PROBOND resin (Invitrogen). Transformed cells are selected on media containing blasticidin.

Spodoptera frugiperda (Sf9) insect cells are infected with recombinant Autographica californica nuclear polyhedrosis virus (baculovirus). The polyhedrin gene is replaced with the cDNA by homologous recombination and the polyhedrin promoter drives cDNA transcription. The protein is synthesized as a fusion protein with 6xHis which enables purification as described above. Purified protein is used in the following activity and to make antibodies

XII Production of Antibodies

STEAPRP is purified using polyacrylamide gel electrophoresis and used to immunize mice or rabbits. Antibodies are produced using the protocols below. Alternatively, the amino acid sequence of STEAPRP is analyzed using LASERGENE software (DNASTAR) to determine regions of high antigenicity. An antigenic epitope, usually found near the C-terminus or in a hydrophilic region is selected, synthesized, and used to raise antibodies. Typically, epitopes of about 15 residues in length are produced using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc-chemistry and coupled to KLH (Sigma-Aldrich) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester to increase antigenicity.

Rabbits are immunized with the epitope-KLH complex in complete Freund's adjuvant. Immunizations are repeated at intervals thereafter in incomplete Freund's adjuvant. After a minimum of seven weeks for mouse or twelve weeks for rabbit, antisera are drawn and tested for antipeptide activity. Testing involves binding the peptide to plastic, blocking with 1% bovine serum albumin, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG. Methods well known in the art are used to determine antibody titer and the amount of complex formation.

XIII Purification of Naturally Occurring Protein Using Specific Antibodies

Naturally occurring or recombinant protein is purified by immunoaffinity chromatography using antibodies which specifically bind the protein. An immunoaffinity column is constructed by covalently coupling the antibody to CNBr-activated SEPHAROSE resin (APB). Media containing the protein is passed over the immunoaffinity column, and the column is washed using high ionic strength buffers in the presence of detergent to allow preferential absorbance of the protein. After coupling, the protein is eluted from the column using a buffer of pH 2-3 or a high concentration of urea or

thiocyanate ion to disrupt antibody/protein binding, and the protein is collected.

XIV Screening Molecules for Specific Binding with the cDNA or Protein

The cDNA, or fragments thereof, or the protein, or portions thereof, are labeled with ³²P-dCTP, Cy3-dCTP, or Cy5-dCTP (APB), or with BIODIPY or FITC (Molecular Probes, Eugene OR), 5 respectively. Libraries of candidate molecules or compounds previously arranged on a substrate are incubated in the presence of labeled cDNA or protein. After incubation under conditions for either a nucleic acid or amino acid sequence, the substrate is washed, and any position on the substrate retaining label, which indicates specific binding or complex formation, is assayed, and the ligand is identified. Data obtained using different concentrations of the nucleic acid or protein are used to 10 calculate affinity between the labeled nucleic acid or protein and the bound molecule.

XV Two-Hybrid Screen

A yeast two-hybrid system, MATCHMAKER LexA Two-Hybrid system (Clontech Laboratories, Palo Alto CA), is used to screen for peptides that bind the protein of the invention. A cDNA encoding the protein is inserted into the multiple cloning site of a pLexA vector, ligated, and 15 transformed into E. coli. cDNA, prepared from mRNA, is inserted into the multiple cloning site of a pB42AD vector, ligated, and transformed into E. coli to construct a cDNA library. The pLexA plasmid and pB42AD-cDNA library constructs are isolated from E. coli and used in a 2:1 ratio to co-transform competent yeast EGY48[p8op-lacZ] cells using a polyethylene glycol/lithium acetate protocol. Transformed yeast cells are plated on synthetic dropout (SD) media lacking histidine (-His), 20 tryptophan (-Trp), and uracil (-Ura), and incubated at 30C until the colonies have grown up and are counted. The colonies are pooled in a minimal volume of 1x TE (pH 7.5), replated on SD/-His/-Leu/-Trp/-Ura media supplemented with 2% galactose (Gal), 1% raffinose (Raf), and 80 mg/ml 5-bromo-4-chloro-3-indolyl β -d-galactopyranoside (X-Gal), and subsequently examined for growth of blue 25 colonies. Interaction between expressed protein and cDNA fusion proteins activates expression of a LEU2 reporter gene in EGY48 and produces colony growth on media lacking leucine (-Leu). Interaction also activates expression of β -galactosidase from the p8op-lacZ reporter construct that produces blue color in colonies grown on X-Gal.

Positive interactions between expressed protein and cDNA fusion proteins are verified by isolating individual positive colonies and growing them in SD/-Trp/-Ura liquid medium for 1 to 2 days 30 at 30C. A sample of the culture is plated on SD/-Trp/-Ura media and incubated at 30C until colonies appear. The sample is replica-plated on SD/-Trp/-Ura and SD/-His/-Trp/-Ura plates. Colonies that grow on SD containing histidine but not on media lacking histidine have lost the pLexA plasmid. Histidine-requiring colonies are grown on SD/Gal/Raf/X-Gal/-Trp/-Ura, and white colonies are 35 isolated and propagated. The pB42AD-cDNA plasmid, which contains a cDNA encoding a protein that physically interacts with the protein, is isolated from the yeast cells and characterized.

XVI STEAPRP Assay

The localization of STEAPRP in the prostate is detected by immunohistochemical analysis as described by Hubert *et al.* (*supra*). Prostate tissue sections (4-mm) are fixed with formalin and embedded in paraffin. Tissues are incubated with anti-STEAPRP antibodies, washed, and then treated 5 with biotinylated rabbit anti-sheep IgG. STEAPRP is visualized with avidin-conjugated horseradish peroxidase (Vector Laboratories, Burlingame CA).

All patents and publications mentioned in the specification are incorporated by reference herein. Various modifications and variations of the described method and system of the invention will 10 be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the field of molecular biology or related fields are intended to be 15 within the scope of the following claims.

Tissue Category	Clone Count	Found in	Abs Abund	Pct Abund
Cardiovascular System	266190	0/68	0	0.0000
Connective Tissue	144645	0/47	0	0.0000
Digestive System	501101	3/148	3	0.0006
Embryonic Structures	106713	0/21	0	0.0000
Endocrine System	225386	2/53	2	0.0009
Exocrine Glands	254635	1/64	1	0.0004
Reproductive, Female	427284	2/106	2	0.0005
Reproductive, Male	448207	28/114	43	0.0096
Germ Cells	38282	0/5	0	0.0000
Hemic and Immune System	680277	2/159	3	0.0004
Liver	109378	1/35	2	0.0018
Musculoskeletal System	159280	2/47	3	0.0019
Nervous System	955753	9/198	12	0.0013
Pancreas	110207	1/24	2	0.0018
Respiratory System	390086	6/93	9	0.0023
Sense Organs	19256	0/8	0	0.0000
Skin	72292	0/15	0	0.0000
Stomatognathic System	12923	0/10	0	0.0000
Unclassified/Mixed	120926	1/13	1	0.0008
Urinary Tract	279062	2/64	2	0.0007
Totals	5321883	60/1292	85	0.0000

TABLE 1

<u>Clone</u>	<u>Library ID</u>	<u>Library Description</u>	<u>Count</u>	<u>Abs</u>	<u>Abund</u>	<u>Pct</u>
3679	PROSNOT19	prostate, AH, mw/adenoCA, M	4	0.1087	4	0.1087
2201	PROSPTT18	prostate tumor, adenoCA, 68M	2	0.0909	2	0.0909
1157	PROETMP06	prostate, PIN, mw/cancer, M	1	0.0864	1	0.0864
3705	PROSNOT26	prostate, mw/adenoCA, 65M	3	0.0810	3	0.0810
3873	PROSDIT01	prostate, AH, mw/adenoCA, 58M	3	0.0775	3	0.0775
4550	PROSTUS20	prostate tumor, adenoCA, 59M, SUB	3	0.0659	3	0.0659
8553	PROSTUT04	prostate tumor, adenoCA, 57M	3	0.0351	3	0.0351
2995	PROSNOT20	prostate, AH, mw/adenoCA, 65M	1	0.0334	1	0.0334
3268	PROSTUT21	prostate tumor, adenoCA, 61M	1	0.0306	1	0.0306
4087	PROSTUS19	prostate tumor, adenoCA, 59M, SUB	1	0.0245	1	0.0245
7138	PROSTUT12	prostate tumor, adenoCA, 65M	1	0.0140	1	0.0140
8829	PROSNOT06	prostate, AH, mw/adenoCA, 57M	1	0.0113	1	0.0113

TABLE 2

mean log ₂ DE (Cy5/Cy3)	CV %	Cy3	Cy5
1.82	0	Human, PPEC Cells, Starved 24hr	Human, LNCaP Line, Starved 24hr, CA
3	0	Human, PPEC Cells, Starved 24hr	Human, LNCaP Line, Starved 24hr, CA
1.86	0	Human, PPEC Cells, Starved 24hr	Human, LNCaP Line, Rich media 24hr, CA
3.02	0	Human, PPEC Cells, Starved 24hr	Human, LNCaP Line, Rich media 24hr, CA
2.5	12.45	Human, PPEC Cells	Human, LNCaP Line, CA

TABLE 3

What is claimed is:

1. An isolated cDNA comprising a nucleic acid sequence encoding a protein having the amino acid sequence of SEQ ID NO:1, or the complement thereof.
2. An isolated cDNA comprising a nucleic acid sequence selected from:
 - 5 a) SEQ ID NO:2 or the complement thereof;
 - b) a fragment of SEQ ID NO:2 selected from SEQ ID NOs:3-9 or the complement thereof; and
 - c) a variant of SEQ ID NO:2 selected from SEQ ID NO:10 or the complement thereof.
3. A composition comprising the cDNA or the complement of the cDNA of claim 1 and a labeling moiety.
- 10 4. A vector comprising the cDNA of claim 1.
5. A host cell comprising the vector of claim 4.
6. A method for using a cDNA to produce a protein, the method comprising:
 - a) culturing the host cell of claim 5 under conditions for protein expression; and
 - b) recovering the protein from the host cell culture.
- 15 7. A method for using a cDNA to detect expression of a nucleic acid in a sample comprising:
 - a) hybridizing the composition of claim 3 to nucleic acids of the sample, thereby forming hybridization complexes; and
 - b) comparing hybridization complex formation with a standard, wherein the comparison indicates expression of the cDNA in the sample.
- 20 8. The method of claim 7 further comprising amplifying the nucleic acids of the sample prior to hybridization.
9. The method of claim 7 wherein the composition is attached to a substrate.
10. The method of claim 7 wherein the cDNA is differentially expressed when compared with a standard and is diagnostic of prostate hyperplasia or prostate cancer.
- 25 11. A method of using a cDNA to screen a plurality of molecules or compounds, the method comprising:
 - a) combining the cDNA of claim 1 with a plurality of molecules or compounds under conditions to allow specific binding; and
 - b) detecting specific binding, thereby identifying a molecule or compound which specifically binds the cDNA.
- 30 12. The method of claim 11 wherein the molecules or compounds are selected from DNA molecules, RNA molecules, peptide nucleic acids, artificial chromosome constructions, peptides, transcription factors, repressors, and regulatory molecules.
13. A purified protein or a portion thereof produced by the method of claim 6 and selected from:
 - 35 a) an amino acid sequence of SEQ ID NO:1;

- b) an antigenic epitope of SEQ ID NO:1; and
- c) a biologically active portion of SEQ ID NO:1.

14. A composition comprising the protein of claim 13 and a pharmaceutical carrier.

15. A method for using a protein to screen a plurality of molecules or compounds to identify at least 5 one ligand, the method comprising:

- a) combining the protein of claim 13 with the molecules or compounds under conditions to allow specific binding; and
- b) detecting specific binding, thereby identifying a ligand which specifically binds the protein.

16. The method of claim 15 wherein the molecules or compounds are selected from DNA molecules, 10 RNA molecules, peptide nucleic acids, peptides, proteins, mimetics, agonists, antagonists, antibodies, immunoglobulins, inhibitors, and drugs.

17. A method of using a protein to prepare and purify antibodies comprising:

- a) immunizing a animal with the protein of claim 15 under conditions to elicit an antibody response;
- 15 b) isolating animal antibodies;
- c) attaching the protein to a substrate;
- d) contacting the substrate with isolated antibodies under conditions to allow specific binding to the protein;
- e) dissociating the antibodies from the protein, thereby obtaining purified antibodies.

20 18. An antibody produced by the method of claim 17.

19. A method for using an antibody to diagnose conditions or diseases associated with expression of a protein, the method comprising:

- a) combining the antibody of claim 18 with a sample, thereby forming antibody:protein complexes; and
- 25 b) comparing complex formation with a standard, wherein the comparison indicates expression of the protein in the sample.

20. The method of claim 19 wherein expression is diagnostic of prostate hyperplasia or prostate cancer.

21. A method for preparing a monoclonal antibody with the specificity of the antibody of claim 18 30 comprising:

- a) immunizing a animal with a protein of SEQ ID NO:1 under conditions to elicit an antibody response;
- b) isolating antibody-producing cells from the animal;
- c) fusing the antibody-producing cells with immortalized cells in culture to form monoclonal 35 antibody producing hybridoma cells;

- d) culturing the hybridoma cells; and
- e) isolating monoclonal antibodies from culture.

22. A monoclonal antibody produced by the method of claim 21.

23. A method for using an antibody to immunopurify a protein comprising:

5 a) attaching an antibody of claim 18 to a substrate,

 b) exposing the antibody to a sample containing protein under conditions to allow antibody:protein complexes to form,

 c) dissociating the protein from the complex, and

 d) collecting the purified protein.

10 24. A composition comprising an antibody of claim 18 and a labeling moiety.

25. A composition comprising an antibody of claim 18 and a pharmaceutical agent.

9	GG GGA AGC AGC TGG AGT GCG ACC GCC ACG GCA GCC ACC CTC CAA CCG CCA GTC	18	27	36	45	54
63	GGA GGT GCA GTC CGT AGG CCC TGG CCC CCG GGT GGG CCC T ^r TG GGG AGT CGG CGC	72	81	90	99	108
117	CGC TCC CGA GGA GCT GCA AGG CTC GCC CCT GCC CGG CGT GGA GGG CGC GGG GGG	126	135	144	153	162
171	CGC GGA GAA AGT GAA GAG AGG AAA TTG GAA AAT TGT GAG TGG ACC TTC TGA TAC	180	189	198	207	216
225	TGC TCC TCC T ^r TG CGT GGA AAA GGG GAA AGA ACT GCA TGC ATA TTA TTC AGC GTC	234	243	252	261	270
279	CTA TAT TCA AAG GAT ATT CTT GGT GAT CTT GGA AGT GTC CGT ATC ATG GAA TCA	288	297	306	315	324
					M E S	
333	ATC TCT ATG GGA AGC CCT AAG AGC CTT AGT GAA ACT TGT TTA CCT AAT GGC	342	351	360	369	378
I	S M M G S P K S L S E T C L P N G					

FIGURE 1A

ATA	AAT	GGT	ATC	AAA	GAT	GCA	AGG	AAG	GTC	ACT	GTA	GGT	GTG	ATT	GGA	AGT	GGA	432
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	N	G	I	K	D	A	R	K	V	T	V	G	V	I	G	S	G	
441		450			459			468			468			477			486	
GAT	TTT	GCC	AAA	TCC	TTG	ACC	ATT	CGA	CTT	ATT	AGA	TGC	GGC	TAT	CAT	GTG	GTC	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
D	F	A	K	S	L	T	I	R	L	I	R	C	G	Y	H	V	V	
495		504			513			522			522			531			540	
ATA	GGA	AGT	AGA	AAT	CCT	AAG	TTT	GCT	TCT	GAA	TTT	TTT	CCT	CAT	GTG	GTA	GAT	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	G	S	R	N	P	K	F	A	S	E	F	F	P	H	V	V	D	
549		558			567			576			576			585			594	
GTC	ACT	CAT	CAT	GAA	GAT	GCT	CTC	ACA	AAA	ACA	AAT	ATA	ATA	TTT	GTT	GCT	ATA	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V	T	H	H	E	D	A	L	T	K	T	N	I	I	F	V	A	I	
603		612			621			630			630			639			648	
CAC	AGA	GAA	CAT	TAT	ACC	TCC	CTG	TGG	GAC	CTG	AGA	CAT	CTG	CTT	GTG	GGT	AAA	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H	R	E	H	Y	T	S	L	W	D	L	R	H	L	L	V	G	K	

FIGURE 1B

ATC	CTG	ATT	GAT	GTG	AGC	AAT	AAC	ATG	AGG	ATA	AAC	CAG	TAC	CCA	GAA	TCC	AAT	
I	L	I	D	V	S	N	N	M	R	I	N	Q	Y	P	E	S	N	
711		720			729				738				747				756	
GCT	GAA	TAT	TTG	GCT	TCA	TTC	TTC	CCA	GAT	TCT	TCT	TTG	ATT	GTC	AAA	GGA	TTT	AAT
A	E	Y	L	A	S	L	F	P	D	S	L	I	V	K	G	F	N	
765		774			783				792				801				810	
GTT	GTC	TCA	GCT	TGG	GCA	CTT	CAG	TAA	GGA	CCT	AAG	GAT	GCC	AGC	CGG	CAG	GT	T
V	V	S	A	W	A	L	Q	L	G	P	K	D	A	S	R	Q	V	
819		828			837				846				855				864	
TAT	ATA	TGC	AGC	AAC	AAT	ATT	CAA	GCG	CGA	CAA	CAG	GTT	ATT	GAA	CTT	GCC	CGC	
Y	I	C	S	N	N	I	Q	A	R	Q	Q	V	I	E	L	A	R	
873		882			891				900				909				918	
CAG	TTG	AAT	TTC	ATT	CCC	ATT	GAC	TTG	GGA	TCC	TTA	TCA	TCA	GCC	AGA	GAG	ATT	
Q	L	N	F	I	P	I	D	L	G	S	L	S	S	A	R	E	I	

FIGURE 1C

GAA	AAT	TTA	CCC	CTA	CGA	CTC	TTT	ACT	CTC	TGG	AGA	GGG	CCA	GTG	GTG	GTA	GCT	972
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	N	L	P	L	R	L	F	T	L	W	R	G	P	V	V	V	A	-
927	936	945	945	954	954	963	963	963	963	963	963	963	963	963	963	963	963	972
981	990	999	999	1008	1008	1017	1017	1017	1017	1017	1017	1017	1017	1017	1017	1017	1017	1026
ATA	AGC	TTG	GCC	ACA	TTA	TTT	TTC	CTT	TAT	TCC	TTT	GTC	AGA	GAT	GTG	ATT	CAT	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	S	L	A	T	F	F	F	L	Y	S	F	V	R	D	V	I	H	-
1035	1044	1053	1053	1062	1062	1071	1071	1071	1071	1071	1071	1071	1071	1071	1071	1071	1071	1080
CCA	TAT	GCT	AGA	AAC	CAA	CAG	AGT	GAC	TTT	TAC	AAA	ATT	CCT	ATA	GAG	ATT	GTG	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P	Y	A	R	N	Q	Q	S	D	F	Y	K	I	P	I	E	I	V	-
1089	1098	1107	1107	1116	1116	1125	1125	1125	1125	1125	1125	1125	1125	1125	1125	1125	1125	1134
AAT	AAA	ACC	TTA	CCT	ATA	GTT	GCC	ATT	ACT	TTG	CTC	TCC	CTA	GTA	TAC	CTC	GCA	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N	K	T	L	P	I	V	A	I	T	L	L	S	L	V	Y	L	A	-
1143	1152	1161	1161	1170	1170	1179	1179	1179	1179	1179	1179	1179	1179	1179	1179	1179	1179	1188
G	L	L	A	A	A	Y	Q	L	Y	Y	G	T	K	Y	R	R	F	-

FIGURE 1D

CCA	CCT	TGG	TTG	GAA	ACC	TGG	TTA	CAG	TGT	AGA	AAA	CAG	CTT	GGA	TTA	CTA	AGT
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P	P	W	L	E	T	W	L	Q	C	R	K	Q	L	G	L	L	S
1251		1260				1269			1278			1287			1296		
TTT	TTC	GCT	GCT	CAT	GTC	GTC	GCC	TAC	AGC	CTC	TGC	TTA	CCG	ATG	AGA	AGG	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F	F	A	M	V	H	V	A	Y	S	L	C	L	P	M	R	R	-
1305		1314				1323			1332			1341			1350		
TCA	GAG	AGA	TAT	TTC	TTG	TTT	CTC	AAC	ATG	GCT	TAT	CAG	GTT	CAT	GCA	AAT	ATT
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S	E	R	Y	L	F	L	N	M	A	Y	Q	Q	V	H	A	N	I
1359		1368				1377			1386			1395			1404		
GAA	AAC	TCT	TGG	AAT	GAG	GAA	GAA	GTT	TGG	AGA	ATT	GAA	ATG	TAT	ATC	TCC	TTT
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	N	S	W	N	E	E	V	W	R	I	E	M	Y	I	S	F	-
1413		1422				1431			1440			1449			1458		
GGC	ATA	ATG	AGC	CTT	GGC	TTA	CTT	TCC	CTC	CTG	GCA	GTC	ACT	TCT	ATC	CCT	TCA
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G	I	M	S	L	G	L	L	S	L	A	V	T	S	I	P	S	-

FIGURE 1E

GTG	AGC	AAT	GCT	TAA	AAC	TGG	AGA	GAA	TTC	AGT	TTT	ATT	CAG	TCT	ACA	CTT	GGA
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V	S	N	A	L	N	W	R	E	F	S	F	I	Q	S	T	L	G
1467	1476	1485	1494	1503	1512												
TAT	GTC	GCT	CTG	CTC	ATA	AGT	ACT	TTC	CAT	GTG	TTA	ATT	TAT	GGA	TGG	AAA	CGA
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Y	V	A	L	L	I	S	T	F	H	V	L	I	Y	G	W	K	R
1521	1530	1539	1548	1557	1566												
GCT	TTC	GAG	GAA	GAG	TAC	TAC	AGA	TTT	TAT	ACA	CCA	CCA	AAC	TTT	GTT	CTT	GCT
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A	F	E	E	Y	Y	R	F	Y	T	P	P	N	F	V	L	A	
1575	1584	1593	1602	1611	1620												
CTT	GTT	TTG	CCC	TCA	ATT	GTA	ATT	CTG	GGT	AAG	ATT	TTA	TTC	CTT	CCA	TGT	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L	V	L	P	S	I	V	I	L	G	K	I	I	L	F	L	P	C
1629	1638	1647	1656	1665	1674												
ATA	AGC	CGA	AAG	CTA	AAA	CGA	ATT	AAA	AAA	GGC	TGG	GAA	AAG	AGC	CAA	TTT	CTG
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	S	R	K	L	K	R	I	K	G	W	E	K	S	Q	F	L	

FIGURE 1F

GAA	GAA	GGT	ATT	GGA	GGA	ACA	ATT	CCT	CAT	GTC	TCC	CCG	GAG	GTC	ACA	GTA	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
E	E	G	I	G	G	T	I	P	H	V	S	P	E	R	V	T	V
1737	1746	1746	1755	1755	1764	1764	1773	1773	1782	1782	1782	1782	1782	1782	1782	1782	1782
ATG	TGA	TAA	ATG	GTG	TTC	ACA	GCT	GCC	ATA	TAA	AGT	TCT	ACT	CAT	GCC	ATT	---
1791	1800	1809	1809	1818	1818	1827	1827	1836	1836	1836	1836	1836	1836	1836	1836	1836	1836
M																	
1845	1854	1863	1863	1872	1872	1881	1881	1890	1890	1890	1890	1890	1890	1890	1890	1890	1890
ATT	TTC	ATG	ACT	TCT	ACG	TTC	AGT	TAC	AAG	TAT	GCT	GTC	AAA	TTA	TCG	TGG	TGT
GA	3'																

FIGURE 1G

1	M E S	I S M M G S P K S L S E T C L P N G I N G I K D A R K	7492448
1	M E S	- - - - -	R K g6572948
31	V T V G V I G S G D F A K S L T I R L I R C G Y H V V I G S	7492448	
6	- - - - -	- - - - -	g6572948
61	R N P K F A S E F F P H V V D V T H H E D A L T K T N I I F	7492448	
6	- - - - -	- - - - -	g6572948
61	D I T N Q E E	- - - - -	- - - - -
91	V A I H R E H Y T S L W D L R H L L V G K I L I D V S N N M	7492448	
13	- - - - -	- - - - -	R R N L g6572948
13	L W K M K P	- - - - -	- - - - -
121	R I N Q Y P E S N A E Y L A S L I F P D S L I V K G F N V V S	7492448	
23	E E D D Y L H K D T G - E T S M L K R P V L -	- - - - -	g6572948
151	A W A L Q I G P K D A S R Q V Y I C S N N I Q A R Q Q V I E	7492448	
44	- - L H L - H Q T A H A D E F D C P S E L Q H T Q E -	- - - - -	g6572948
181	L A R Q L N F I P I D L G S L S S A R E I E N L P L R L F T	7492448	
67	- - - - -	- - - - -	L F P g6572948
211	L W R G P V V V A I S L A T F F F L Y S F V R D V I H P Y A	7492448	
70	Q W H L P I K I A I T I A S L T F L Y T L L R E V I H P L A	g6572948	

FIGURE 2A

241 R N Q Q S D F Y K I P I E I V N K T L P I V A I T L L S L V 7492448
 100 T S H Q Q F Y K I P I L V I N K V L P M V S I T L L A L V g6572948

271 Y L A G L L A A Y Q L Y G T K Y R R F P P W L E T W L Q 7492448
 130 Y L P G V I A A I V Q L H N G T K Y K K F P H W L D K W M L g6572948

301 C R K Q L G L L S F F F A M V H V A Y S L C L P M R R S E R 7492448
 160 T R K Q F G L L S F F F A V L H A T Y S L S Y P M R R S Y R g6572948

331 Y L F L N M A Y Q Q V H A N I E N S W N E E V W R I E M Y 7492448
 190 Y K L L N W A Y Q Q V Q Q N K E D A W T E H D V W R M E I Y g6572948

361 I S F G I M S L G L I S L L A V T S I P S V S N A I N W R E 7492448
 220 V S L G I V G L A I L A V T S I P S V S D S L T W R E g6572948

391 F S F I Q S T L G Y V A L L I S T F H V L I Y G W K R A F E 7492448
 250 F H Y I Q S K L G T V S L L L G T I H A L I F A W N K W I D g6572948

421 E E Y Y R F Y T P P N F V L A L V L P S I V I L G K I I L F 7492448
 280 I K Q F V W Y T P P T F M I A V F L P I V V L I F K S I L F g6572948

FIGURE 2B

481	H V S P E R V - T V M	7492448
329	K I N K T E I C S Q L	96572948
451	L P C I S R K L K R T K K G W E K S Q F L E E G I G G T I P	7492448
310	L P C L R K K I L K I R H G W E D	96572948

FIGURE 2C

<110> INCYTE GENOMICS, INC.
LAL, Preeti G.
FARIS, Mary
CHEN, Huei-Mei
ISON, Craig H.

<120> STEAP-RELATED PROTEIN

<130> PC-0037 PCT

<140> To Be Assigned
<141> Herewith

<151> 09/802,520
<152> 2001-03-09

<160> 11

<170> PERL Program

<210> 1
<211> 490
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7492448CD1

<400> 1
Met Glu Ser Ile Ser Met Met Gly Ser Pro Lys Ser Leu Ser Glu
1 5 10 15
Thr Cys Leu Pro Asn Gly Ile Asn Gly Ile Lys Asp Ala Arg Lys
20 25 30
Val Thr Val Gly Val Ile Gly Ser Gly Asp Phe Ala Lys Ser Leu
35 40 45
Thr Ile Arg Leu Ile Arg Cys Gly Tyr His Val Val Ile Gly Ser
50 55 60
Arg Asn Pro Lys Phe Ala Ser Glu Phe Phe Pro His Val Val Asp
65 70 75
Val Thr His His Glu Asp Ala Leu Thr Lys Thr Asn Ile Ile Phe
80 85 90
Val Ala Ile His Arg Glu His Tyr Thr Ser Leu Trp Asp Leu Arg
95 100 105
His Leu Leu Val Gly Lys Ile Leu Ile Asp Val Ser Asn Asn Met
110 115 120
Arg Ile Asn Gln Tyr Pro Glu Ser Asn Ala Glu Tyr Leu Ala Ser
125 130 135
Leu Phe Pro Asp Ser Leu Ile Val Lys Gly Phe Asn Val Val Ser
140 145 150
Ala Trp Ala Leu Gln Leu Gly Pro Lys Asp Ala Ser Arg Gln Val
155 160 165
Tyr Ile Cys Ser Asn Asn Ile Gln Ala Arg Gln Gln Val Ile Glu
170 175 180
Leu Ala Arg Gln Leu Asn Phe Ile Pro Ile Asp Leu Gly Ser Leu
185 190 195
Ser Ser Ala Arg Glu Ile Glu Asn Leu Pro Leu Arg Leu Phe Thr
200 205 210
Leu Trp Arg Gly Pro Val Val Val Ala Ile Ser Leu Ala Thr Phe
215 220 225
Phe Phe Leu Tyr Ser Phe Val Arg Asp Val Ile His Pro Tyr Ala
230 235 240
Arg Asn Gln Gln Ser Asp Phe Tyr Lys Ile Pro Ile Glu Ile Val

245	250	255
Asn Lys Thr Leu Pro Ile Val Ala Ile	Thr Leu Leu Ser Leu	Val
260	265	270
Tyr Leu Ala Gly Leu Leu Ala Ala Ala	Tyr Gln Leu Tyr Tyr	Gly
275	280	285
Thr Lys Tyr Arg Arg Phe Pro Pro Trp	Leu Glu Thr Trp Leu	Gln
290	295	300
Cys Arg Lys Gln Leu Gly Leu Leu Ser	Phe Phe Phe Ala Met	Val
305	310	315
His Val Ala Tyr Ser Leu Cys Leu Pro	Met Arg Arg Ser Glu	Arg
320	325	330
Tyr Leu Phe Leu Asn Met Ala Tyr Gln	Gln Val His Ala Asn	Ile
335	340	345
Glu Asn Ser Trp Asn Glu Glu Val	Trp Arg Ile Glu Met	Tyr
350	355	360
Ile Ser Phe Gly Ile Met Ser Leu Gly	Leu Leu Ser Leu Leu	Ala
365	370	375
Val Thr Ser Ile Pro Ser Val Ser Asn	Ala Leu Asn Trp Arg	Glu
380	385	390
Phe Ser Phe Ile Gln Ser Thr Leu Gly	Tyr Val Ala Leu Leu	Ile
395	400	405
Ser Thr Phe His Val Leu Ile Tyr Gly	Trp Lys Arg Ala Phe	Glu
410	415	420
Glu Glu Tyr Tyr Arg Phe Tyr Thr Pro	Pro Asn Phe Val Leu	Ala
425	430	435
Leu Val Leu Pro Ser Ile Val Ile Leu	Gly Lys Ile Ile Leu	Phe
440	445	450
Leu Pro Cys Ile Ser Arg Lys Leu Lys	Arg Ile Lys Lys Gly	Trp
455	460	465
Glu Lys Ser Gln Phe Leu Glu Glu Gly	Ile Gly Gly Thr Ile	Pro
470	475	480
His Val Ser Pro Glu Arg Val Thr Val	Met	
485	490	

<210> 2

<211> 1891

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7492448CB1

<400> 2

```

ggggaaagcag ctggagtgcg accggcacgg cagccacccct gcaaccgcca gtcggaggtg 60
cagtcgttag gccctggccc cgggtgggc ccttggggag tcggcgccgc tcccgaggag 120
ctgcaaggct cggccctgccc cggcggtggag ggcgcggggg ggcgcggagaa agtgaagaga 180
ggaaatttggaa aaatttggag tggacattct gatactgttc ctcttgcgt ggaaaagggg 240
aaagaactgc atgcataat ttcagcgatcc tatattccaa ggatattttt ggtgatcttg 300
gaagtgtccg tatcatggaa tcaatctcta tgatgggaag ccctaagagc ctttagtgaaa 360
cttggttacc taatggcata aatgttatca aagatgcag gaaggtcact gtaggtgtga 420
ttggaaagtgg agatttgcc aaatcccttga ccattcgact tattagatgc ggctatcatg 480
tggtcatagg aagtagaaat cctaagtttgc cttctgaatt ttttcctcat gtggtagatg 540
tcactccatca tgaagatgtc ctcacaaaaaa caaatataat atttgttgct atacacagag 600
aacattatac ctcctgtgg gacctgagac atctgctgtt gggtaaaatc ctgattgtg 660
tgagcaataa catgaggata aaccagtacc cagaatccaa tgctgaatat ttggcttcat 720
tattcccaga ttctttgatt gtcaaaaggat ttaatgttgct ctcagcttgg gcacttcagt 780
taggacctaa ggatgccagc cggcagggtt atatatgcag caacaatatt caagcgcgac 840
aacagggttat tgaacttggc cgccagttga atttcattcc cattgacttg ggatccttat 900
catcagccag agagattgaa aatttacccc tacgactctt tactctctgg agagggccag 960
tggtggtagc tataagcttg gccacatttt ttttccttta ttctttgtc agagatgtga
1020
ttcatccata tgctagaaac caacagagtg acttttacaa aattcctata gagattgtga

```


tgttgtata cacagagaac attatacctc cctgtggac ctgagacatc tgcttgaaa 240
 taaaatccctg attgatgtga gcaataacat gaggataaac cagtagccag aatccatgc 300
 tgaatatttg gcttcattat tcccaagatc tttgattgtc aaaggattta atgttgctc 360
 agcttggca cttcagttag gacctaagga tgccagccgg caggttata tatgcagcaa 420
 caatattcaa gcgcgacaaac aggttattga acttgccgc cagttgaatt tcattccat 480
 tgacttggga tcc 493

<210> 5
 <211> 403
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4647117F6

<220>
 <221> unsure
 <222> 316, 321, 339
 <223> a, t, c, g, or other

<400> 5
 cccagattct ttgattgtca aaggatttaa tgttgtctca gcttggcac ttcagttagg 60
 acctaaggat gccagccggc aggttataat atgcagcaac aatattcaag cgcgacaaca 120
 ggttattgaa cttgcccggc agttgaattt cattccatt gacttggat ccttatcatc 180
 agccagagag attgaaaatt taccctacg actctttact ctctggagag ggccagtggt 240
 ggttagctata agcttggcca cattttttt cctttattcc tttgtcagag atgtgattca 300
 tccatatgct agaaaancaaac ngagtgactt ttacaaacnt tctatagaga ttgtgaataa 360
 aaccttacct atagttgcca ttactttgct ccccttagta tac 403

<210> 6
 <211> 560
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7004364H1

<400> 6
 acattttttt tccttgcgtgc ctttgcaga gatgtgattc atccatatgc tagaaaccaa 60
 cagagtgact tttacaaaat tcctatagag attgtgaata aaaccttacc tatagttgcc 120
 attacatttgc tctccctagt atacctcgca ggtcttcgtt cagctgctta tcaactttat 180
 tacggcacca agtataaggag atttccaccc tgggtggaaa cctggttaca gtgttagaaaa 240
 cagcttggat tactaagttt tatcttcgtt atggtccatg ttgcctacag cctctgctta 300
 ccgatgagaa ggtcagagag atatttgttt ctcaacatgg cttatcagca ggttcatgca 360
 aatattgaaa actcttggaa tgaggaagaa gtttggagaa ttgaaatgtt tatctccttt 420
 ggcataatga gccttggctt acttccctc ctggcagtca cttctatccc ttcaagtggc 480
 aatgcattaa actggagaga attcagttt attcagtcata cacttggata tgtcgctctg 540
 ctcataagta ctttccatgt 560

<210> 7
 <211> 265
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 70351677D1

<400> 7
 ctcagtctgg gtatctgaa actgcaaaag atccagaatt acaattgagg gcaaaaacaag 60
 agcaagaaca aagtttggtg gtgtataaaa tctgttagtac tcttcctcaa aagctcggtt 120

ccatccataa attaaaacat ggaaagtact tatgagcaga gcgacatatac caagttaga 180
 ctgaataaaa ctgaattctc tccagttaa agcattgctc actgaaggga tagaagtgac 240
 tgccaggagg gaaaagtaagc caagg 265

<210> 8
 <211> 204
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4108079H1

<220>
 <221> unsure
 <222> 45, 83, 132
 <223> a, t, c, g, or other

<400> 8
 cagagtttat acaccaccaa acttggttct tgctcggttt ttgcnctcag gtgttaattct 60
 ggggaagatt gttttattcc ttngtgtata aggcgaaagc taaaacgaat taagaaaaggc 120
 tggggaaaga gnccgatttc tggagaagg tctgggaggg acaattcgca tgtcgccccg 180
 gagagggtca cagtaatggg atga 204

<210> 9
 <211> 265
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4669848H1

<400> 9
 ccggagaggg tcacagtaat gtgatgataa atgggtttca cagctgcccataaaagttct 60
 actcatgcca ttattttat gacttctacg ttcaagtata agtatgtgtt caaattatcg 120
 tgggtgaaa cttgttaat gagatttcaa ctgacttagt gatagagttt tcttcaagtt 180
 aattttcaca aatgtcatgt ttgccaatat gaattttct agtcaacata ttattgtaat 240
 ttaggtatgt tttgttttgtt tttgc 265

<210> 10
 <211> 525
 <212> DNA
 <213> Rattus norvegicus

<220>
 <221> misc_feature
 <223> Incyte ID No: 702819778T1

<400> 10
 gggatgtgta atgttctcta tggatagcca cgaatattat atttgtcttc gttaaagcgt 60
 cttcatggtg ggtgacgtct accacatgag gaaaaaaactc agacgcgaac ttaggatttc 120
 tgcttccgat gaccacgtga tagccgcacc tgataagccg aatggtcaga gacttggcaa 180
 aatccccact tcctatcacc cccacggta ccttccctgc gtctttgata ccgtttatgc 240
 cattaggcaa aaacgtctcc agggtcttag ggcttcccat catagagatg gattccatgg 300
 tagagactct tctaagatca ccagaatgc cctgggaatc ttaaggtgta gcttctcact 360
 cagaggagct ggagggaggg tccttcggcg ctgctggact ctgaaactgc ctacgtgtag 420
 tgaggagggc tcggcgcccc tcctctcccg gcccacggcg cagcgcccgccg ccgtggctcc 480
 ctcgcgccaa gggcccgcccg agctcccgccg cctacggagt gctcc 525

<210> 11
 <211> 339
 <212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: g6572948

<400> 11

Met	Glu	Ser	Arg	Lys	Asp	Ile	Thr	Asn	Gln	Glu	Glu	Leu	Trp	Lys	
1				5					10					15	
Met	Lys	Pro	Arg	Arg	Arg	Asn	Leu	Glu	Glu	Asp	Asp	Tyr	Leu	His	Lys
					20				25					30	
Asp	Thr	Gly	Glu	Thr	Ser	Met	Leu	Lys	Arg	Pro	Val	Leu	Leu	His	
					35				40					45	
Leu	His	Gln	Thr	Ala	His	Ala	Asp	Glu	Phe	Asp	Cys	Pro	Ser	Glu	
					50				55					60	
Leu	Gln	His	Thr	Gln	Glu	Leu	Phe	Pro	Gln	Trp	His	Leu	Pro	Ile	
					65				70					75	
Lys	Ile	Ala	Ala	Ile	Ile	Ala	Ser	Leu	Thr	Phe	Leu	Tyr	Thr	Leu	
					80				85					90	
Leu	Arg	Glu	Val	Ile	His	Pro	Leu	Ala	Thr	Ser	His	Gln	Gln	Tyr	
					95				100					105	
Phe	Tyr	Lys	Ile	Pro	Ile	Leu	Val	Ile	Asn	Lys	Val	Leu	Pro	Met	
					110				115					120	
Val	Ser	Ile	Thr	Leu	Leu	Ala	Leu	Val	Tyr	Leu	Pro	Gly	Val	Ile	
					125				130					135	
Ala	Ala	Ile	Val	Gln	Leu	His	Asn	Gly	Thr	Lys	Tyr	Lys	Lys	Phe	
					140				145					150	
Pro	His	Trp	Leu	Asp	Lys	Trp	Met	Leu	Thr	Arg	Lys	Gln	Phe	Gly	
					155				160					165	
Leu	Leu	Ser	Phe	Phe	Phe	Ala	Val	Leu	His	Ala	Ile	Tyr	Ser	Leu	
					170				175					180	
Ser	Tyr	Pro	Met	Arg	Arg	Ser	Tyr	Arg	Tyr	Lys	Leu	Leu	Asn	Trp	
					185				190					195	
Ala	Tyr	Gln	Gln	Val	Gln	Gln	Asn	Lys	Glu	Asp	Ala	Trp	Ile	Glu	
					200				205					210	
His	Asp	Val	Trp	Arg	Met	Glu	Ile	Tyr	Val	Ser	Leu	Gly	Ile	Val	
					215				220					225	
Gly	Leu	Ala	Ile	Leu	Ala	Leu	Leu	Ala	Val	Thr	Ser	Ile	Pro	Ser	
					230				235					240	
Val	Ser	Asp	Ser	Leu	Thr	Trp	Arg	Glu	Phe	His	Tyr	Ile	Gln	Ser	
					245				250					255	
Lys	Leu	Gly	Ile	Val	Ser	Leu	Leu	Leu	Gly	Thr	Ile	His	Ala	Leu	
					260				265					270	
Ile	Phe	Ala	Trp	Asn	Lys	Trp	Ile	Asp	Ile	Lys	Gln	Phe	Val	Trp	
					275				280					285	
Tyr	Thr	Pro	Pro	Thr	Phe	Met	Ile	Ala	Val	Phe	Leu	Pro	Ile	Val	
					290				295					300	

Val	Leu	Ile	Phe	Lys	Ser	Ile	Leu	Phe	Leu	Pro	Cys	Leu	Arg	Lys
				305				310					315	
Lys	Ile	Leu	Lys	Ile	Arg	His	Gly	Trp	Glu	Asp	Val	Thr	Lys	Ile
					320				325					330
Asn	Lys	Thr	Glu	Ile	Cys	Ser	Gln	Leu						
					335									

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07053

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/02, 21/04; C12P 21/06; C12N 5/00
US CL : 536/23.1, 23.5; 435/69.1, 325

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1, 23.5; 435/69.1, 325

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MPSRCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,329,503 B1 (AFAR et al) 11 December 2001, SEQ ID NO:8.	1
P		3-6
X	Database Genbank, Accession No. AC002064, GATTUNG et al. Gene Sequence. 09 May 1997. See entire document.	2

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

16 June 2002 (16.06.2002)

Date of mailing of the international search report

22 AUG 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

ANTHONY CAPUTA

Telephone No. 703-308-0916

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07053

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-6

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/07053

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I. Claims 1-6, drawn to a nucleic acid sequence of SEQ ID NO:2, or the complement thereof, or a nucleic acid sequence encoding SEQ ID NO:1, fragments selected from SEQ ID NO:3-9, and a variant of SEQ ID NO:10, a vector comprising a nucleic acid encoding SEQ ID NO:1, a host cell comprising said vector, and a method for producing a protein, classified in class 536, subclass 23.1. Group II. Claims 7, 9, drawn to a method for detecting the expression of a nucleic acid, using hybridization, classified in class 435, subclass 6.

Group III. Claim 8, drawn to a method for detecting the expression of a nucleic acid, using amplification, classified in class 435, subclass .

Group IV. Claim 10, drawn to a method for detecting prostate hyperplasia or prostate cancer, classified in class 435, subclass 6.

Group V. Claims 11-12, drawn to a method for screening compounds that bind specifically to a nucleic acid encoding SEQ ID NO:1, classified in class 435, subclass 6.

Group VI. Claims 13-14, drawn to a protein of SEQ ID NO:1 or fragments thereof, classified in class 530, subclass 350.

Group VII. Claims 15-16, drawn to a method for screening compounds that specifically binds to SEQ ID NO:1, classified in class 435, subclass 7.1.

Group VIII. Claims 18,21-24, drawn to an antibody, classified in class 530, subclass 387.1.

Group IX. Claims 19-20, drawn to a method for detecting prostate hyperplasia or prostate cancer, using an antibody, classified in class 435, subclass 7.1.

In addition, upon election of group I, further election of the following species is required:

Full length sequence of SEQ ID NO:2, or fragments or a variant thereof

Upon election of fragments of SEQ ID NO:2, further election of the following species is required:

SEQ ID Nos: 3-9.

The inventions listed as Groups I-IX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

According to PCT Rule 13.2, unity of invention exists only when the shared same or corresponding technical feature is a contribution over the prior art. The inventions listed as groups I-IX do not relate to a single general inventive concept because they lack the same or corresponding special technical feature. The technical feature of group I, a complement of SEQ ID NO:2, which is shown to be the same as the inherent complement of a nucleic acid molecule of SEQ ID NO:8 taught by PN=6329503, lacks novelty and does not make a contribution over the prior art.

The species do not share the same technical feature, because they have different structure.